STIC Search Report
Biotech-Chem Library

STIC Database Tracking Number: 165274

TO: Janet Epps-Ford **Art Unit:** 1653 **Location:** REM-2C05&2C18 **Serial Number:** 09/915543 **Friday, June 10, 2005**	**From:** Beverly Shears **Location:** Biotech-Chem Library REM 1A54 **Phone:** 571-272-2528 beverly.shears@uspto.gov

Search Notes

Protein Sequence Searches – February 2005

All of the sequence databases on ABSS have recently been updated.

- Please note that the curators of the UniProt database have purged some temporary accession numbers from the most recent version of UniProt. These sequences have been assigned new permanent accession numbers. The new UniProt record may not contain the previous temporary accession number.

- If you encounter an accession number from an older search run against UniProt (results file extension **.rup**) that can no longer be found in the database, the permanent record with the new accession number can be found by searching the old accession number in the UniProt Protein Archive database (uniPARC) at:

  http://www.pir.uniprot.org/database/archive.shtml

If you have any questions regarding this information or your results, please contact any STIC searcher.

STIC-Biotech/ChemLib 155274

From: Epps-Ford, Janet
Sent: Thursday, June 02, 2005 3:48 PM
To: STIC-Biotech/ChemLib
Subject: Amino acid sequence search

Application No. 09/915,543

Isolated polypeptides comprising:

a) a peptide consisting of amino acids 177 to 204 of SEQ ID NO: 15;

or aa 1426 ARE

b) a peptide consisting of amino acids 349 to 384 of SEQ ID NO: 15;

wherein said isolated polypeptide does not comprise both of (a) and (b).

Please search all pending and commercial amino acid databases.

n

Thanks,
Janet L. Epps-Ford, Ph.D.
Art Unit 1635
Mailbox: Remsen 2C18
Office: Remsen 2C05
Phone: 571-272-0757
Fax: 571-273-0757

MEy

****
STAFF USE ONLY

Searcher:_____
Searcher Phone: 2-
Date Searcher Picked up:____
Date Completed:_____
Searcher Prep/Rev. Time:___
Online Time:_____

****
Type of Search

NA#:_____ AA#:_____
Interference:____ SPDI:___
S/L:_____ Oligomer:____
Encode/Transl:_____
Structure#:_____ Text:___
Inventor:____ Litigation:___

****
Vendors and cost where applicable
STN:_____
DIALOG:_____
QUESTEL/ORBIT:_____
LEXIS/NEXIS:_____
SEQUENCE SYSTEM:_____
WWW/Internet:_____
Other(Specify):_____

Date completed: _____
Searcher: Beverly e 2528
Terminal time: _____
Elapsed time: _____
CPU time: _____
Total time: _____
Number of Searches: _____
Number of Databases: _____

Search Site
_____ STIC
_____ CM-1
_____ Pre-S

Type of Search
_____ N.A. Sequence
_____ A.A. Sequence
_____ Structure
_____ Bibliographic

Vendors
_____ IG
_____ STN
_____ Dialog
_____ APS
_____ Geninfo
_____ SDC
_____ DARC/Questel
✓ Other CGN

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: June 8, 2005, 03:09:59 ; Search time 27.5625 Seconds
(without alignments)
125.671 Million cell updates/sec

Title: US-09-915-543-15_COPY_349_384
Perfect score: 183
Sequence: 1 DGLSQEQLEHRERSLQTLRDIQRMLFPDEKEFTGAQ 36

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_79:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	57	31.1	1034	2	AB0551	exonuclease SbcC [
2	56	30.6	584	2	S51882	topoisomerase I-re
3	56	30.6	859	2	T46372	hypothetical prote
4	55	30.1	287	2	AE2895	transcription regu
5	55	30.1	295	2	H97670	hex regulon repres
6	55	30.1	643	2	B59436	Rho GTPase activat
7	54.5	29.8	1171	2	S14065	phytochrome B - ri
8	53.5	29.2	266	2	AE1124	1-pyrroline-5-carb
9	53.5	29.2	4957	2	T03455	ALR protein - huma
10	53.5	29.2	5262	2	T03454	ALR protein - huma
11	53	29.0	227	2	B70438	hypothetical prote
12	53	29.0	376	1	E69957	gamma-D-glutamyl-L
13	53	29.0	818	2	S62790	mismatch DNA recog
14	52.5	28.7	242	2	B70366	hypothetical prote
15	52.5	28.7	705	2	T24343	hypothetical prote
16	52.5	28.7	1039	2	T14802	phytochrome B - so
17	52	28.4	332	2	B47017	probable transcrip
18	52	28.4	332	2	AD2541	transcription init
19	52	28.4	572	2	D82984	pyruvate dehydroge
20	52	28.4	1009	2	S61174	hypothetical prote
21	51	27.9	102	2	AH0216	conserved hypothet
22	51	27.9	237	2	A85901	probable alpha hel
23	51	27.9	237	2	A49940	probable alpha hel
24	51	27.9	237	2	E91056	probable alpha hel
25	51	27.9	329	2	D96834	hypothetical prote
26	51	27.9	477	2	T18801	hypothetical prote
27	51	27.9	518	2	G86454	CDS protein F9L11.
28	51	27.9	899	1	GNMVMM	pol polyprotein -
29	51	27.9	1047	2	C85535	ATP-dependent dsDN
30	51	27.9	1047	2	G90684	ATP-dependent dsDN
31	51	27.9	1161	2	G81186	conserved hypothet
32	51	27.9	1161	2	G81915	hypothetical prote
33	51	27.9	1464	2	S58984	development protei
34	50.5	27.6	835	2	AD2441	endopeptidase Clp
35	50	27.3	273	2	H69337	conserved hypothet
36	50	27.3	275	2	H69843	hypothetical prote
37	50	27.3	319	2	S49771	hypothetical prote
38	50	27.3	788	2	S67595	hypothetical prote
39	50	27.3	1162	2	D83454	conserved hypothet
40	50	27.3	1236	2	B36329	hypothetical prote
41	49.5	27.0	302	1	TPCHTC	troponin T, cardia
42	49.5	27.0	336	2	S72858	hypothetical prote
43	49.5	27.0	830	2	T18860	hypothetical prote
44	49.5	27.0	2101	2	A42184	nuclear mitotic ap
45	49	26.8	310	2	A84142	L-lactate dehydrog

ALIGNMENTS

RESULT 1
AB0551
exonuclease SbcC [imported] - Salmonella enterica subsp. enterica serovar Typhi (strain
C;Species: Salmonella enterica subsp. enterica serovar Typhi
A;Note: this species has also been called Salmonella typhi
C;Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 18-Nov-2002
C;Accession: AB0551
R;Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher,
th, T.; Connerton, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar,
, S.; Moule, S.; O'Gaora, P.
Nature 413, 848-852, 2001
A;Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.;
A;Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serov
A;Reference number: AB0502; MUID:21534947; PMID:11677608
A;Accession: AB0551
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1034 <PAR>
A;Cross-references: GB:AL513382; PIDN:CAD08850.1; PID:g16501663; GSPDB:GN00176
C;Genetics:
A;Gene: STY0429
C;Superfamily: sbcC protein

Query Match 31.1%; Score 57; DB 2; Length 1034;
Best Local Similarity 42.9%; Pred. No. 30;
Matches 12; Conservative 6; Mismatches 10; Indels 0; Gaps 0;

```
Qy          3 LSQEQLEHRERSLQTLRDIQRMLFPDEK 30
              |: |||:  | |||  | |  :: |  |::
Db        213 LADEQLQQLEASLQALTDEEKRLLADQQ 240
```

RESULT 2
S51882
topoisomerase I-related protein TRF4 - yeast (Saccharomyces cerevisiae)
N;Alternate names: protein HRC584; protein O0716; protein YOL115w
C;Species: Saccharomyces cerevisiae
C;Date: 05-May-1995 #sequence_revision 03-Aug-1995 #text_change 09-Jul-2004
C;Accession: S51882; S59158; S58774; S66811
R;Vandenbol, M.; Durand, P.; Portetelle, D.; Hilger, F.
submitted to the EMBL Data Library, January 1995
A;Description: Sequence analysis of a 44kb DNA fragment of yeast chromosome XV including
and a Delta.
A;Reference number: S51848
A;Accession: S51882
A;Molecule type: DNA
A;Residues: 1-584 <VAN>
A;Cross-references: UNIPROT:P53632; EMBL:Z48149; NID:g663234; PID:g663237
R;Vandenbol, M.; Durand, P.; Portetelle, D.; Hilger, F.
Yeast 11, 1069-1075, 1995
A;Title: Sequence analysis of a 44 kb DNA fragment of yeast chromosome XV including the
a delta element.

```
A;Reference number: S59156; MUID:96076631; PMID:7502582
A;Accession: S59158
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-584 <VAW>
A;Cross-references: EMBL:Z48149; NID:g663234; PIDN:CAA88145.1; PID:g663237
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, January 1995
R;Sadoff, B.U.; Heath-Pagliuso, S.; Castano, I.B.; Zhu, Y.; Kieff, F.S.; Christman, M.F.
Genetics 141, 465-479, 1995
A;Title: Isolation of mutants of Saccharomyces cerevisiae requiring DNA topoisomerase I.
A;Reference number: S58774; MUID:96109595; PMID:8647385
A;Accession: S58774
A;Molecule type: DNA
A;Residues: 1-584 <SAD>
A;Cross-references: EMBL:U31355; NID:g950225; PIDN:AAC49091.1; PID:g950226
R;Durand, P.; Hilger, F.; Portetelle, D.; Vandenbol, M.
submitted to the Protein Sequence Database, July 1996
A;Reference number: S66791
A;Accession: S66811
A;Molecule type: DNA
A;Residues: 1-584 <DUR>
A;Cross-references: EMBL:Z74857; NID:g1419986; PID:e251905; PID:g1419987; MIPS:YOL115w
A;Experimental source: strain S288C
C;Genetics:
A;Gene: SGD:TRF4
A;Cross-references: SGD:S0005475; MIPS:YOL115w
A;Map position: 15L
C;Keywords: nucleus

  Query Match             30.6%;  Score 56;  DB 2;  Length 584;
  Best Local Similarity   36.0%;  Pred. No. 22;
  Matches    9;  Conservative   10;  Mismatches    6;  Indels    0;  Gaps    0;

Qy          4 SQEQLEHRERSLQTLRDIQRMLFPD 28
              |:|::| | ::: |:|:  : |:||
Db        193 SREEIEIRNQTISTIREAVKQLWPD 217


RESULT 3
T46372
hypothetical protein DKFZp434P1818.1 - human (fragment)
C;Species: Homo sapiens (man)
C;Date: 04-Feb-2000 #sequence_revision 04-Feb-2000 #text_change 09-Jul-2004
C;Accession: T46372
R;Ottenwaelder, B.; Obermaier, B.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.
submitted to the Protein Sequence Database, January 2000
A;Reference number: Z23031
A;Accession: T46372
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-859 <AAA>
A;Cross-references: UNIPROT:Q9NT51; EMBL:AL137528
A;Experimental source: adult testis; clone DKFZp434P1818
C;Genetics:
A;Note: DKFZp434P1818.1

  Query Match             30.6%;  Score 56;  DB 2;  Length 859;
  Best Local Similarity   33.3%;  Pred. No. 33;
  Matches   10;  Conservative    8;  Mismatches   12;  Indels    0;  Gaps    0;

Qy          5 QEQLEHRERSLQTLRDIQRMLFPDEKEFTG 34
              : |  |:|   | |:   ::| ||:::| |
Db        317 ENQRSHQELISQLLQSYMKLLLPDDEKFHG 346


RESULT 4
AE2895
transcription regulator, RpiR family Atu2598 [imported] - Agrobacterium tumefaciens (str
C;Species: Agrobacterium tumefaciens
C;Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 09-Jul-2004
C;Accession: AE2895
R;Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, L
erage, G.; Gillet, W.; Grant, C.; Guenthner, D.; Kutyavin, T.; Levy, R.; Li, M.; McClell
; Karp, P.; Romero, P.; Zhang, S.
Science 294, 2317-2323, 2001
A;Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,
ster, E.W.
A;Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
A;Reference number: AB2577; MUID:21608550; PMID:11743193
A;Accession: AE2895
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-287 <KUR>
A;Cross-references: UNIPROT:Q8UCA0; GB:AE008688; PIDN:AAL43579.1; PID:g17741095; GSPDB:G
A;Experimental source: strain C58 (Dupont)
C;Genetics:
A;Gene: Atu2598
A;Map position: circular chromosome
C;Superfamily: hypothetical protein ybbH

  Query Match             30.1%;  Score 55;  DB 2;  Length 287;
  Best Local Similarity   40.0%;  Pred. No. 13;
  Matches   12;  Conservative    7;  Mismatches    7;  Indels    4;  Gaps    1;

Qy          9 EHRERSLQTLRDIQRMLF----PDEKEFTG 34
              : |:||: ||| |:: |     ||:|:  |
Db        257 QQRQRSMVTLRHIKQQLVEHRDPDDKQLLG 286


RESULT 5
H97670
hex regulon repressor [imported] - Agrobacterium tumefaciens (strain C58, Cereon)
C;Species: Agrobacterium tumefaciens
C;Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 09-Jul-2004
C;Accession: H97670
R;Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Goldman,
 A.; Liu, F.; Wollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz, B.;
Science 294, 2323-2328, 2001
A;Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tume
A;Reference number: A97359; MUID:21608551; PMID:11743194
A;Accession: H97670
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-295 <KUR>
A;Cross-references: UNIPROT:Q8UCA0; GB:AE007869; PIDN:AAK88321.1; PID:g15157797; GSPDB:GN
C;Genetics:
A;Gene: AGR_C_4707
A;Map position: circular chromosome
C;Superfamily: hypothetical protein ybbH

  Query Match             30.1%;  Score 55;  DB 2;  Length 295;
  Best Local Similarity   40.0%;  Pred. No. 14;
  Matches   12;  Conservative    7;  Mismatches    7;  Indels    4;  Gaps    1;

Qy          9 EHRERSLQTLRDIQRMLF----PDEKEFTG 34
              : |:||: ||| |:: |     ||:|:  |
Db        265 QQRQRSMVTLRHIKQQLVEHRDPDDKQLLG 294


RESULT 6
B59436
Rho GTPase activating protein RhoGAP8 - human
C;Species: Homo sapiens (man)
C;Date: 03-Jun-2002 #sequence_revision 03-Jun-2002 #text_change 09-Jul-2004
C;Accession: B59436
R;Goward, M.E.; Huckle, E.J.
submitted to GenBank, April 2000
A;Reference number: B59436
A;Accession: B59436
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-643 <GOW>
A;Cross-references: UNIPROT:Q9NSG0; GB:CAB90248; PID:g7711011; PIDN:CAB90248.1

  Query Match             30.1%;  Score 55;  DB 2;  Length 643;
  Best Local Similarity   52.2%;  Pred. No. 32;
```

```
  Matches   12;  Conservative    4;  Mismatches    7;  Indels    0;  Gaps    0;

Qy          2 GLSQEQLEHRERSLQTLRDIQRM 24
              ||  | |  |  |:||:|:|||:
Db        434 GLRTEGLFRRSASVQTVREIQRL 456


RESULT 7
S14065
phytochrome B - rice
C;Species: Oryza sativa (rice)
C;Date: 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change 09-Jul-2004
C;Accession: S14065
R;Dehesh, K.; Tepperman, J.; Christensen, A.H.; Quail, P.H.
Mol. Gen. Genet. 225, 305-313, 1991
A;Title: phyB is evolutionarily conserved and constitutively expressed in rice seedling
A;Reference number: S14065; MUID:91172131; PMID:2005872
A;Accession: S14065
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1171 <DEH>
A;Cross-references: UNIPROT:P25764; GB:X57563; NID:g6469490; PIDN:CAA40795.2; PID:g64694
C;Genetics:
A;Gene: phyB
C;Superfamily: phytochrome; phytochrome homology
C;Keywords: chromoprotein; photoreceptor; phytochromobilin; transcription regulation
F;103-623/Domain: phytochrome homology <PHYT>
F;364/Binding site: phytochromobilin (Cys) (covalent) #status predicted

  Query Match             29.8%;  Score 54.5;  DB 2;  Length 1171;
  Best Local Similarity   48.1%;  Pred. No. 72;
  Matches   13;  Conservative    4;  Mismatches    5;  Indels    5;  Gaps    1;

Qy          3 LSQEQLEHRERSLQTLRDIQRMLFPDE 29
              :||  :: ||| || :|||     |||
Db       1019 VSQVMIQLRERDLQLIRDI-----PDE 1040


RESULT 8
AE1124
1-pyrroline-5-carboxylate reductase (ProC) homolog lmo0396 [imported] - Listeria monocyt
C;Species: Listeria monocytogenes
C;Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 09-Jul-2004
C;Accession: AE1124
R;Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloecker
.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H.
D.; Jones, L.M.; Karst, U.
Science 294, 849-852, 2001
A;Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Ma
ok, C.; Schlueter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland,
A;Title: Comparative genomics of Listeria species.
A;Reference number: AB1077; MUID:21537279; PMID:11679669
A;Accession: AE1124
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-266 <GLA>
A;Cross-references: UNIPROT:Q8Y9X2; GB:NC_003210; PIDN:CAC98475.1; PID:g16409774; GSPDB:
A;Experimental source: strain EGD-e
C;Genetics:
A;Gene: lmo0396
C;Superfamily: pyrroline-5-carboxylate reductase

  Query Match             29.2%;  Score 53.5;  DB 2;  Length 266;
  Best Local Similarity   42.9%;  Pred. No. 19;
  Matches   15;  Conservative    6;  Mismatches   13;  Indels    1;  Gaps    1;

Qy          2 GLSQEQLEHRERSLQTLRDIQRMLFPDEKEFTGAQ 36
              ||:|  |  ||  :   |:::: | | | |||| |
Db         19 GLAQANLVKREEIIVGGRNLEK-LKPLEAEFTGLQ 52


RESULT 9
T03455
ALR protein - human
C;Species: Homo sapiens (man)
C;Date: 24-Mar-1999 #sequence_revision 24-Mar-1999 #text_change 09-Jul-2004
C;Accession: T03455
R;Prasad, R.; Zhadanov, A.B.; Sedkov, Y.; Bullrich, F.; Druck, T.; Rallapalli, R.; Yano,
Oncogene 15, 549-560, 1997
A;Title: Structure and expression pattern of human ALR, a novel gene with strong homolog
A;Reference number: Z14954; MUID:97388474; PMID:9247308
A;Accession: T03455
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-4957 <PRA>
A;Cross-references: UNIPROT:O14686; EMBL:AF010404; NID:g2358286; PIDN:AAC51735.1; PID:g2
C;Genetics:
A;Gene: ALR
A;Map position: 12
C;Superfamily: acute lymphoblastic leukemia protein, ALR type
C;Keywords: alternative splicing

  Query Match             29.2%;  Score 53.5;  DB 2;  Length 4957;
  Best Local Similarity   42.5%;  Pred. No. 4.7e+02;
  Matches   17;  Conservative    4;  Mismatches   12;  Indels    7;  Gaps    2;

Qy          2 GLSQEQLEHRERSLQTLRD------IQRMLFPDEKEFTGA 35
              |||| :|| ::|  | ||:      |||     |||   |
Db       2091 GLSQTELE-KQRQRQRLRELLIRQQIQRNTLRQEKETAAA 2129


RESULT 10
T03454
ALR protein - human
C;Species: Homo sapiens (man)
C;Date: 24-Mar-1999 #sequence_revision 24-Mar-1999 #text_change 09-Jul-2004
C;Accession: T03454
R;Prasad, R.; Zhadanov, A.B.; Sedkov, Y.; Bullrich, F.; Druck, T.; Rallapalli, R.; Yano,
Oncogene 15, 549-560, 1997
A;Title: Structure and expression pattern of human ALR, a novel gene with strong homolog
A;Reference number: Z14954; MUID:97388474; PMID:9247308
A;Accession: T03454
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-5262 <PRA>
A;Cross-references: UNIPROT:O14686; EMBL:AF010403; NID:g2358284; PIDN:AAC51734.1; PID:g2
C;Genetics:
A;Gene: ALR
A;Map position: 12
C;Superfamily: acute lymphoblastic leukemia protein, ALR type
C;Keywords: alternative splicing

  Query Match             29.2%;  Score 53.5;  DB 2;  Length 5262;
  Best Local Similarity   42.5%;  Pred. No. 5e+02;
  Matches   17;  Conservative    4;  Mismatches   12;  Indels    7;  Gaps    2;

Qy          2 GLSQEQLEHRERSLQTLRD------IQRMLFPDEKEFTGA 35
              |||| :|| ::|  | ||:      |||     |||   |
Db       2396 GLSQTELE-KQRQRQRLRELLIRQQIQRNTLRQEKETAAA 2434


RESULT 11
B70438
hypothetical protein aq_1596 - Aquifex aeolicus
C;Species: Aquifex aeolicus
C;Date: 08-May-1998 #sequence_revision 08-May-1998 #text_change 09-Jul-2004
C;Accession: B70438
R;Deckert, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.; Ov
V.
Nature 392, 353-358, 1998
A;Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.
A;Reference number: A70300; MUID:98196666; PMID:9537320
A;Accession: B70438
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-227 <AQF>
```

```
A;Cross-references: UNIPROT:O67532; GB:AE000747; NID:g2983944; PIDN:AAC07503.1; PID:g298
A;Experimental source: strain VF5
C;Genetics:
A;Gene: aq_1596

  Query Match             29.0%;  Score 53;  DB 2;  Length 227;
  Best Local Similarity   41.4%;  Pred. No. 19;
  Matches   12;  Conservative    7;  Mismatches    8;  Indels    2;  Gaps    1;

Qy          5 QEQLEHRERSLQTLRDIQRML--FPDEKE 31
              :| || ||: |:|:|  :  |  | :||:
Db         33 KELLEEREKLLETIRSYEEKLDSFEEEKQ 61


RESULT 12
E69957
gamma-D-glutamyl-L-diamino acid endopeptid homolog yqgT - Bacillus subtilis
C;Species: Bacillus subtilis
C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 09-Jul-2004
C;Accession: E69957
R;Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berter
C.; Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Cho
A.; Ehrlich, S.D.; Emmerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.
Nature 390, 249-256, 1997
A;Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Galler
iech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.
Koetter, P.; Koningstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois,
A;Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maueel
y, M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle
 Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scanlon,
A;Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Seror
akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama,
T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K
A;Authors: Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Danchin, A.
A;Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.
A;Reference number: A69580; MUID:98044033; PMID:9384377
A;Accession: E69957
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-376 <KUN>
A;Cross-references: UNIPROT:P54497; GB:Z99116; GB:AL009126; NID:g2634723; PIDN:CAB14414.
A;Experimental source: strain 168
C;Genetics:
A;Gene: yqgT
C;Superfamily: endopeptidase I

  Query Match             29.0%;  Score 53;  DB 1;  Length 376;
  Best Local Similarity   52.6%;  Pred. No. 33;
  Matches   10;  Conservative    4;  Mismatches    5;  Indels    0;  Gaps    0;

Qy          9 EHRERSLQTLRDIQRMLFP 27
              ||  : |||::||:| | |
Db         50 EHSGKELQTIQDIKRFLQP 68


RESULT 13
S62790
mismatch DNA recognition protein mutS [validated] - Thermus aquaticus (fragment)
C;Species: Thermus aquaticus
C;Date: 19-Mar-1997 #sequence_revision 25-Apr-1997 #text_change 12-Jul-2004
C;Accession: S62790
R;Takamatsu, S.; Kato, R.; Kuramitsu, S.
Nucleic Acids Res. 24, 640-647, 1996
A;Title: Mismatch DNA recognition protein from an extremely thermophilic bacterium, Ther
A;Reference number: S62790; MUID:96177563; PMID:8604304
A;Accession: S62790
A;Status: preliminary; nucleic acid sequence not shown
A;Molecule type: DNA
A;Residues: 1-818 <TAK>
A;Cross-references: EMBL:D63810
A;Note: the source is designated as Thermus thermophilus
C;Genetics:
A;Gene: mutS
C;Function:
A;Description: implicated in DNA mismatch repair; binds to DNA and specifically recogniz
atched DNA [validated, MUID:96177563]

  Query Match             29.0%;  Score 53;  DB 2;  Length 818;
  Best Local Similarity   42.3%;  Pred. No. 76;
  Matches   11;  Conservative    5;  Mismatches   10;  Indels    0;  Gaps    0;

Qy          6 EQLEHRERSLQTLRDIQRMLFPDEKE 31
              |::    | :|||:| ||   |: ||
Db        465 EKVPQEYRPVQTLKDRQRYTLPEMKE 490


RESULT 14
B70366
hypothetical protein aq_755 - Aquifex aeolicus
C;Species: Aquifex aeolicus
C;Date: 08-May-1998 #sequence_revision 08-May-1998 #text_change 09-Jul-2004
C;Accession: B70366
R;Deckert, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.; Ove
V.
Nature 392, 353-358, 1998
A;Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.
A;Reference number: A70300; MUID:98196666; PMID:9537320
A;Accession: B70366
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-242 <AQF>
A;Cross-references: UNIPROT:O66957; GB:AE000706; NID:g2983327; PIDN:AAC06923.1; PID:g298
A;Experimental source: strain VF5
C;Genetics:
A;Gene: aq_755
C;Superfamily: Aquifex aeolicus hypothetical protein aq_755

  Query Match             28.7%;  Score 52.5;  DB 2;  Length 242;
  Best Local Similarity   45.8%;  Pred. No. 24;
  Matches   11;  Conservative    7;  Mismatches    5;  Indels    1;  Gaps    1;

Qy         10 HRERSLQT-LRDIQRMLFPDEKEF 32
              |:  ||:  :|:|::|:|  ||||
Db         60 HKRTSLRKFVREIEKMVFEAEKEF 83


RESULT 15
T24343
hypothetical protein T02B5.1 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
C;Accession: T24343
R;McMurray, A.
submitted to the EMBL Data Library, October 1996
A;Reference number: Z19878
A;Accession: T24343
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-705 <WIL>
A;Cross-references: UNIPROT:O01302; EMBL:Z81112; PIDN:CAB03272.1; GSPDB:GN00023; CESP:T0
A;Experimental source: clone T02B5
C;Genetics:
A;Gene: CESP:T02B5.1
A;Map position: 5
A;Introns: 22/2; 88/2; 117/3; 185/1; 221/3; 280/3; 349/2; 427/2; 532/2; 637/1
C;Superfamily: cholinesterase; cholinesterase homology

  Query Match             28.7%;  Score 52.5;  DB 2;  Length 705;
  Best Local Similarity   36.8%;  Pred. No. 75;
  Matches   14;  Conservative    7;  Mismatches   12;  Indels    5;  Gaps    1;

Qy          3 LSQEQLEH-----RERSLQTLRDIQRMLFPDEKEFTGA 35
              ||:||:|:     |::| | : | |  |  :    | ||
Db        273 LSEEQVENTYSCLRKKSAQQILDAQLWLLQNSTYFLGA 310
```

```
RESULT 16
T14802
phytochrome B - sorghum (fragment)
C;Species: Sorghum bicolor (sorghum)
C;Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 05-May-2000
C;Accession: T14802
R;Childs, K.L.; Miller, F.R.; Cordonnier-Pratt, M.M.; Pratt, L.H.; Morgan, P.W.; Mullet,
submitted to the EMBL Data Library, April 1996
A;Description: The Sorghum bicolor photoperiod sensitivity gene, Ma3, encodes a phytochr
A;Reference number: Z18185
A;Accession: T14802
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1039 <CHI>
A;Cross-references: EMBL:U56730; NID:g1800216; PID:g1800217
A;Experimental source: cultivar 58M
C;Genetics:
A;Gene: PHYB
A;Note: Intron positions not resolved (incomplete sequence)
C;Superfamily: phytochrome; phytochrome homology
C;Keywords: chromoprotein; photoreceptor; phytochromobilin
F;233/Binding site: phytochromobilin (Cys) (covalent) #status predicted

  Query Match             28.7%;  Score 52.5;  DB 2;  Length 1039;
  Best Local Similarity   51.9%;  Pred. No. 1.2e+02;
  Matches   14;  Conservative    2;  Mismatches    6;  Indels    5;  Gaps    1;

Qy          3 LSQEQLEHRERSLQTLRDIQRMLFPDE 29
              :||  |  ||| || :|||     |||
Db        887 VSQAMLLLRERDLQLIRDI-----PDE 908


RESULT 17
B47017
probable transcription initiation factor sigma SigB - Anabaena sp.
C;Species: Anabaena sp.
C;Date: 21-Sep-1993 #sequence_revision 18-Nov-1994 #text_change 18-Jun-1999
C;Accession: B47017
R;Brahamsha, B.; Haselkorn, R.
J. Bacteriol. 174, 7273-7282, 1992
A;Title: Identification of multiple RNA polymerase sigma factor homologs in the cyanobac
A;Reference number: A47017; MUID:93054341; PMID:1385387
A;Accession: B47017
A;Status: preliminary
A;Molecule type: nucleic acid
A;Residues: 1-332 <BRA>
A;Cross-references: GB:M95760; NID:g142111; PIDN:AAA22046.1; PID:g142112
A;Experimental source: PCC 7120
A;Note: sequence extracted from NCBI backbone (NCBIN:118034, NCBIP:118036)
C;Superfamily: transcription initiation factor sigma katF; transcription initiation fact
C;Keywords: DNA binding; sigma factor; transcription initiation
F;103-328/Domain: transcription initiation factor sigma katF homology <KTF>

  Query Match             28.4%;  Score 52;  DB 2;  Length 332;
  Best Local Similarity   35.5%;  Pred. No. 39;
  Matches   11;  Conservative    6;  Mismatches   14;  Indels    0;  Gaps    0;

Qy          1 DGLSQEQLEHRERSLQTLRDIQRMLFPDEKE 31
              ||:| |:   ||   | : ::   | | :||
Db        248 DGMSPERYAERELLYQDIHNLLAKLTPQQKE 278


RESULT 18
AD2541
transcription initiation factor sigma sigB [imported] - Nostoc sp. (strain PCC 7120) pla
C;Species: Nostoc sp. PCC 7120
A;Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C;Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Jul-2004
C;Accession: AD2541
R;Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi
 Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S
DNA Res. 8, 205-213, 2001
A;Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Ana
A;Reference number: AB1807; MUID:21595285; PMID:11759840
A;Accession: AD2541
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-332 <KUR>
A;Cross-references: UNIPROT:Q03065; GB:AP003602; PIDN:BAB77258.1; PID:g17134700; GSPDB:G
A;Experimental source: strain PCC 7120
C;Genetics:
A;Gene: sigB
A;Genome: plasmid
C;Superfamily: transcription initiation factor sigma katF; transcription initiation fact
C;Keywords: transcription initiation

  Query Match             28.4%;  Score 52;  DB 2;  Length 332;
  Best Local Similarity   35.5%;  Pred. No. 39;
  Matches   11;  Conservative    6;  Mismatches   14;  Indels    0;  Gaps    0;

Qy          1 DGLSQEQLEHRERSLQTLRDIQRMLFPDEKE 31
              ||:| |:   ||   | : ::   | | :||
Db        248 DGMSPERYAERELLYQDIHNLLAKLTPQQKE 278


RESULT 19
D82984
pyruvate dehydrogenase (cytochrome) PA5297 [imported] - Pseudomonas aeruginosa (strain P
C;Species: Pseudomonas aeruginosa
C;Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 16-Aug-2004
C;Accession: D82984
R;Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrener, P.; Hickey, M.J.; Br
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim,
.; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A;Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho
A;Reference number: A82950; MUID:20437337; PMID:10984043
A;Accession: D82984
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-572 <STO>
A;Cross-references: UNIPROT:Q9HTQ7; GB:AE004942; GB:AE004091; NID:g9951607; PIDN:AAG0868
A;Experimental source: strain PAO1
C;Genetics:
A;Gene: poxB; PA5297
C;Superfamily: Acetolactate synthase, large subunit/pyruvate oxidase; thiamin pyrophosph

  Query Match             28.4%;  Score 52;  DB 2;  Length 572;
  Best Local Similarity   84.6%;  Pred. No. 70;
  Matches   11;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          6 EQLEHRERSLQTL 18
              | |||||||:||
Db        333 ECLEHRERSLRTL 345


RESULT 20
S61174
hypothetical protein YDR379w - yeast (Saccharomyces cerevisiae)
N;Alternate names: hypothetical protein D9481.4
C;Species: Saccharomyces cerevisiae
C;Date: 23-Feb-1996 #sequence_revision 01-Mar-1996 #text_change 15-Mar-2004
C;Accession: S61174
R;Ding, H.
submitted to the EMBL Data Library, June 1995
A;Description: The sequence of S. cerevisiae cosmid 9481.
A;Reference number: S61159
A;Accession: S61174
A;Molecule type: DNA
A;Residues: 1-1009 <DIN>
A;Cross-references: EMBL:U28373; NID:g849184; PIDN:AAB64815.1; PID:g849200; MIPS:YDR379w
A;Experimental source: strain S288C (AB972)
C;Genetics:
A;Gene: SGD:RGA2
A;Cross-references: SGD:S0002787; MIPS:YDR379w
A;Map position: 4R
```

```
F;13-66/Domain: LIM metal-binding repeat homology <LIM2>

  Query Match             28.4%;  Score 52;  DB 2;  Length 1009;
  Best Local Similarity   34.1%;  Pred. No. 1.3e+02;
  Matches   15;  Conservative    2;  Mismatches   13;  Indels   14;  Gaps    1;

Qy          3 LSQEQLEHRERSLQTLRDIQRMLFPD----------EKEF 32
              || |   ||  |||| |    :|  |              ||:|
Db        575 LSSESAHHRSSSLQTSRSTNALLEDDSTKVDATDESATSLEKDF 618


RESULT 21
AH0216
conserved hypothetical protein YPO1778 [imported] - Yersinia pestis (strain CO92)
C;Species: Yersinia pestis
C;Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 09-Jul-2004
C;Accession: AH0216
R;Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prentice, M.B.
deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.;
il, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrell,
Nature 413, 523-527, 2001
A;Title: Genome sequence of Yersinia pestis, the causative agent of plague.
A;Reference number: AB0001; MUID:21470413; PMID:11586360
A;Accession: AH0216
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-102 <KUR>
A;Cross-references: UNIPROT:Q8ZFD6; GB:AL590842; PIDN:CAC90596.1; PID:g15979803; GSPDB:G
C;Genetics:
A;Gene: YPO1778
C;Superfamily: uncharacterized conserved protein

  Query Match             27.9%;  Score 51;  DB 2;  Length 102;
  Best Local Similarity   31.2%;  Pred. No. 14;
  Matches   10;  Conservative    6;  Mismatches   16;  Indels    0;  Gaps    0;

Qy          1 DGLSQEQLEHRERSLQTLRDIQRMLFPDEKEF 32
              ||||:   |    ||  |: : : ::|    |
Db         63 DGLSERHAEQENMSLDELKKVIKAIYPGLDRF 94


RESULT 22
A85901
probable alpha helix protein yfhG [imported] - Escherichia coli (strain O157:H7, substra
C;Species: Escherichia coli
C;Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001
C;Accession: A85901
R;Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
iller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,
Nature 409, 529-533, 2001
A;Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A;Reference number: A85480; MUID:21074935; PMID:11206551
A;Accession: A85901
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-237 <STO>
A;Cross-references: GB:AE005174; NID:g12516965; PIDN:AAG57669.1; GSPDB:GN00145; UWGP:Z38
A;Experimental source: strain O157:H7, substrain EDL933
C;Genetics:
A;Gene: yfhG

  Query Match             27.9%;  Score 51;  DB 2;  Length 237;
  Best Local Similarity   52.4%;  Pred. No. 36;
  Matches   11;  Conservative    3;  Mismatches    7;  Indels    0;  Gaps    0;

Qy          5 QEQLEHRERSLQTLRDIQRML 25
              |:|||   | |: | ||:| |
Db        185 QQQLELTTRKLENLTDIERQL 205


RESULT 23
A49940
probable alpha helix protein [imported] - Escherichia coli (strain K-12)
C;Species: Escherichia coli
C;Date: 13-Sep-1994 #sequence_revision 18-Nov-1994 #text_change 09-Jul-2004
C;Accession: A49940; B65033
R;Liu, J.; Magasanik, B.
J. Bacteriol. 175, 7441-7449, 1993
A;Title: The glnB region of the Escherichia coli chromosome.
A;Reference number: A49940; MUID:94042920; PMID:8226691
A;Accession: A49940
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-237 <LIU>
A;Cross-references: UNIPROT:P37328; GB:S67014; NID:g455660; PIDN:AAB28777.1; PID:g455661
A;Note: sequence extracted from NCBI backbone (NCBIN:139878, NCBIP:139880)
R;Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Col
.A.; Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A;Title: The complete genome sequence of Escherichia coli K-12.
A;Reference number: A64720; MUID:97426617; PMID:9278503
A;Accession: B65033
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-237 <BLAT>
A;Cross-references: GB:AE000341; GB:U00096; NID:g1788899; PIDN:AAC75608.1; PID:g1788906;
A;Experimental source: strain K-12, substrain MG1655
C;Genetics:
A;Gene: yfhG

  Query Match             27.9%;  Score 51;  DB 2;  Length 237;
  Best Local Similarity   52.4%;  Pred. No. 36;
  Matches   11;  Conservative    3;  Mismatches    7;  Indels    0;  Gaps    0;

Qy          5 QEQLEHRERSLQTLRDIQRML 25
              |:|||   | |: | ||:| |
Db        185 QQQLELTTRKLENLTDIERQL 205


RESULT 24
E91056
probable alpha helix protein [imported] - Escherichia coli (strain O157:H7, substrain RIM
C;Species: Escherichia coli
C;Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 09-Jul-2004
C;Accession: E91056
R;Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.,
gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001
A;Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and genom
A;Reference number: A99629; MUID:21156231; PMID:11258796
A;Accession: E91056
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-237 <HAY>
A;Cross-references: UNIPROT:P37328; GB:BA000007; PIDN:BAB36844.1; PID:g13362892; GSPDB:GN
A;Experimental source: strain O157:H7, substrain RIMD 0509952
C;Genetics:
A;Gene: ECs3421

  Query Match             27.9%;  Score 51;  DB 2;  Length 237;
  Best Local Similarity   52.4%;  Pred. No. 36;
  Matches   11;  Conservative    3;  Mismatches    7;  Indels    0;  Gaps    0;

Qy          5 QEQLEHRERSLQTLRDIQRML 25
              |:|||   | |: | ||:| |
Db        185 QQQLELTTRKLENLTDIERQL 205


RESULT 25
D96834
hypothetical protein F5I6.4 [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 09-Jul-2004
C;Accession: D96834
R;Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
 Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;
```

ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A;Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziali,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A;Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A;Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A;Reference number: A86141; MUID:21016719; PMID:11130712
A;Accession: D96834
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-329 <STO>
A;Cross-references: UNIPROT:Q9C975; GB:AE005173; NID:g6751704; PIDN:AAF27686.1; GSPDB:GN
C;Genetics:
A;Gene: F5I6.4
A;Map position: 1

Query Match 27.9%; Score 51; DB 2; Length 329;
Best Local Similarity 37.5%; Pred. No. 51;
Matches 9; Conservative 7; Mismatches 8; Indels 0; Gaps 0;

```
Qy          2 GLSQEQLEHRERSLQTLRDIQRML 25
              |||  ::|||:|    :|:  |::
Db        272 GLSSRRVEHRKRRGNCIREFHRVM 295
```

RESULT 26
T18801
hypothetical protein ZK131.11 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
C;Accession: T18801; T27746
R;Percy, C.
submitted to the EMBL Data Library, March 1997
A;Reference number: Z19024
A;Accession: T18801
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-477 <WIL>
A;Cross-references: UNIPROT:O62022; EMBL:Z93373; PIDN:CAB07552.1; GSPDB:GN00020; CESP:ZK
A;Experimental source: clone C01B9
R;Steward, C.
submitted to the EMBL Data Library, December 1996
A;Reference number: Z20413
A;Accession: T27746
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-477 <WI2>
A;Cross-references: EMBL:Z83245; PIDN:CAB05840.1; GSPDB:GN00020; CESP:ZK131.11
A;Experimental source: clone ZK131
C;Genetics:
A;Gene: CESP:ZK131.11
A;Map position: 2
A;Introns: 20/2; 49/2; 113/1; 169/1; 260/2; 375/3; 432/3
C;Superfamily: Caenorhabditis elegans hypothetical protein ZK131.11

Query Match 27.9%; Score 51; DB 2; Length 477;
Best Local Similarity 50.0%; Pred. No. 77;
Matches 11; Conservative 3; Mismatches 8; Indels 0; Gaps 0;

```
Qy          3 LSQEQLEHRERSLQTLRDIQRM 24
              | :|  | |||  | | |:||:
Db        428 LKREMKEMRERDAQILSDLQRV 449
```

RESULT 27
G86454
CDS protein F9L11.17 [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 09-Jul-2004
C;Accession: G86454
R;Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;
ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A;Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziali,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A;Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon, I
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A;Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A;Reference number: A86141; MUID:21016719; PMID:11130712
A;Accession: G86454
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-518 <STO>
A;Cross-references: UNIPROT:Q9MAP1; GB:AE005172; NID:g6910585; PIDN:AAF31290.1; GSPDB:GN
C;Genetics:
A;Gene: F9L11.17
A;Map position: 1
C;Superfamily: Arabidopsis thaliana hypothetical protein At2g23160

Query Match 27.9%; Score 51; DB 2; Length 518;
Best Local Similarity 34.1%; Pred. No. 84;
Matches 14; Conservative 9; Mismatches 10; Indels 8; Gaps 3;

```
Qy          1 DGLSQEQLEHRERSLQTLRDI-----QRMLFPDEKEFTGAQ 36
              ||  :: : |::|   | : |     ||:|  |: ||||  :
Db        442 DG-DEDTVAHQQRDHHTFKSISKFKAQRLL--DDDEFTGVK 479
```

RESULT 28
GNMVMM
pol polyprotein - mouse mammary tumor virus
C;Species: mouse mammary tumor virus, MMTV
C;Date: 31-Mar-1989 #sequence_revision 31-Mar-1989 #text_change 09-Jul-2004
C;Accession: C26795
R;Moore, R.; Dixon, M.; Smith, R.; Peters, G.; Dickson, C.
J. Virol. 61, 480-490, 1987
A;Title: Complete nucleotide sequence of a milk-transmitted mouse mammary tumor virus: t
A;Reference number: A93030; MUID:87112944; PMID:3027377
A;Accession: C26795
A;Molecule type: DNA
A;Residues: 1-899 <MOO>
A;Cross-references: UNIPROT:P03365; EMBL:M15122
C;Genetics:
A;Gene: pol
C;Superfamily: pol polyprotein
C;Keywords: polyprotein; reverse transcriptase

Query Match 27.9%; Score 51; DB 1; Length 899;
Best Local Similarity 40.0%; Pred. No. 1.5e+02;
Matches 10; Conservative 7; Mismatches 8; Indels 0; Gaps 0;

```
Qy          1 DGLSQEQLEHRERSLQTLRDIQRML 25
              | :| ::|: |   |:|| | |::|
Db        240 DSVSYQKLQIRTDKLRTLNDFQKLL 264
```

RESULT 29
C85535
ATP-dependent dsDNA exonuclease [imported] - Escherichia coli (strain O157:H7, substrain
C;Species: Escherichia coli
C;Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 09-Jul-2004
C;Accession: C85535
R;Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
iller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,
Nature 409, 529-533, 2001
A;Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A;Reference number: A85480; MUID:21074935; PMID:11206551
A;Accession: C85535
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1047 <STO>
A;Cross-references: UNIPROT:Q8XEJ6; GB:AE005174; NID:g12513240; PIDN:AAG54743.1; GSPDB:G

```
A;Experimental source: strain O157:H7, substrain EDL933
C;Genetics:
A;Gene: sbcC
C;Superfamily: sbcC protein

  Query Match             27.9%;  Score 51;  DB 2;  Length 1047;
  Best Local Similarity   48.0%;  Pred. No. 1.8e+02;
  Matches   12;  Conservative    2;  Mismatches   11;  Indels    0;  Gaps    0;

Qy          5 QEQLEHRERSLQTLRDIQRMLFPDE 29
              |:|| | |: |  |  |  ||  ||
Db        386 QQQLTHAEQKLNALAAITLMLTADE 410


RESULT 30
G90684
ATP-dependent dsDNA exonuclease [imported] - Escherichia coli (strain O157:H7, substrain
C;Species: Escherichia coli
C;Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 09-Jul-2004
C;Accession: G90684
R;Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.
gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001
A;Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and genc
A;Reference number: A99629; MUID:21156231; PMID:11258796
A;Accession: G90684
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1047 <HAY>
A;Cross-references: UNIPROT:Q8XEJ6; GB:BA000007; PIDN:BAB33870.1; PID:g13359904; GSPDB:G
A;Experimental source: strain O157:H7, substrain RIMD 0509952
C;Genetics:
A;Gene: ECs0447
C;Superfamily: sbcC protein

  Query Match             27.9%;  Score 51;  DB 2;  Length 1047;
  Best Local Similarity   48.0%;  Pred. No. 1.8e+02;
  Matches   12;  Conservative    2;  Mismatches   11;  Indels    0;  Gaps    0;

Qy          5 QEQLEHRERSLQTLRDIQRMLFPDE 29
              |:|| | |: |  |  |  ||  ||
Db        386 QQQLTHAEQKLNALAAITLMLTADE 410


RESULT 31
G81186
conserved hypothetical protein NMB0545 [imported] - Neisseria meningitidis (strain MC58
C;Species: Neisseria meningitidis
C;Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 09-Jul-2004
C;Accession: G81186
R;Tettelin, H.; Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, K.E.; Eisen, J.A
Hickey, E.K.; Haft, D.H.; Salzberg, S.L.; White, O.; Fleischmann, R.D.; Dougherty, B.A.;
ri, H.; Qin, H.; Vamathevan, J.; Gill, J.; Scarlato, V.; Masignani, V.; Pizza, M.
Science 287, 1809-1815, 2000
A;Authors: Grandi, G.; Sun, L.; Smith, H.O.; Fraser, C.M.; Moxon, E.R.; Rappuoli, R.; Ve
A;Title: Complete genome sequence of Neisseria meningitidis serogroup B strain MC58.
A;Reference number: A81000; MUID:20175755; PMID:10710307
A;Accession: G81186
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1161 <TET>
A;Cross-references: UNIPROT:Q9K0P1; GB:AE002410; GB:AE002098; NID:g7225766; PIDN:AAF4097
A;Experimental source: serogroup B, strain MC58
C;Genetics:
A;Gene: NMB0545
C;Superfamily: chromosome segregation protein SMC1

  Query Match             27.9%;  Score 51;  DB 2;  Length 1161;
  Best Local Similarity   33.3%;  Pred. No. 2e+02;
  Matches   14;  Conservative    6;  Mismatches   14;  Indels    8;  Gaps    1;

Qy          3 LSQEQL--------EHRERSLQTLRDIQRMLFPDEKEFTGAQ 36
              | |:||        :| ||  :  :: | :  ||| |   ||
Db        400 LKQQQLAHAEQTIAKHEERKGRLKQENQALNLPDEAETAAAQ 441


RESULT 32
G81915
hypothetical protein NMA0724 [imported] - Neisseria meningitidis (strain Z2491 serogroup
C;Species: Neisseria meningitidis
C;Date: 05-May-2000 #sequence_revision 05-May-2000 #text_change 09-Jul-2004
C;Accession: G81915
R;Parkhill, J.; Achtman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.; Morel
; Holroyd, S.; Jagels, K.; Leather, S.; Moule, S.; Mungall, K.; Quail, M.A.; Rajandream,
Nature 404, 502-506, 2000
A;Title: Complete DNA sequence of a serogroup A strain of Neisseria menigitidis Z2491.
A;Reference number: A81775; MUID:20222556; PMID:10761919
A;Accession: G81915
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1161 <PAR>
A;Cross-references: UNIPROT:Q9JVR9; GB:AL162754; GB:AL157959; NID:g7379424; PIDN:CAB8400
A;Experimental source: serogroup A, strain Z2491
C;Genetics:
A;Gene: NMA0724
C;Superfamily: chromosome segregation protein SMC1

  Query Match             27.9%;  Score 51;  DB 2;  Length 1161;
  Best Local Similarity   33.3%;  Pred. No. 2e+02;
  Matches   14;  Conservative    6;  Mismatches   14;  Indels    8;  Gaps    1;

Qy          3 LSQEQL--------EHRERSLQTLRDIQRMLFPDEKEFTGAQ 36
              | |:||        :| ||  :  :: | :  ||| |   ||
Db        400 LKQQQLAHAEQTVAKHEERKGRLKQENQALNLPDEAETAAAQ 441


RESULT 33
S58984
development protein tolkin (EC 3.4.24.-) - fruit fly (Drosophila melanogaster)
C;Species: Drosophila melanogaster
C;Date: 19-Jul-1996 #sequence_revision 26-Jul-1996 #text_change 09-Jul-2004
C;Accession: S58984
R;Finelli, A.L.; Xie, T.; Bossie, C.A.; Blackman, R.K.; Padgett, R.W.
Genetics 141, 271-281, 1995
A;Title: The tolkin gene is a tolloid/BMP-1 homologue that is essential for Drosophila de
A;Reference number: S58984; MUID:96042912; PMID:8536976
A;Accession: S58984
A;Molecule type: mRNA
A;Residues: 1-1464 <FIN>
A;Cross-references: UNIPROT:Q24132; EMBL:U34777; NID:g1002985; PIDN:AAC47015.1; PID:g1002
A;Note: the authors did not translate the codon for residue 722
C;Genetics:
A;Gene: tolkin
A;Cross-references: FlyBase:FBgn0004885
C;Keywords: hydrolase; metalloproteinase; zinc
F;529-722/Domain: astacin homology <AST>
F;958-993/Domain: EGF homology <EGF>
F;1118-1153/Domain: EGF homology <EGF1>
F;614,618,624,673/Binding site: zinc (His, His, His, Tyr) #status predicted
F;615/Active site: Glu #status predicted

  Query Match             27.9%;  Score 51;  DB 2;  Length 1464;
  Best Local Similarity   33.3%;  Pred. No. 2.6e+02;
  Matches   11;  Conservative    8;  Mismatches   14;  Indels    0;  Gaps    0;

Qy          1 DGLSQEQLEHRERSLQTLRDIQRMLFPDEKEFT 33
              :||  : ||||::    |  |:|:  |   ::|
Db         96 EGLKHQHLEHRQQFDAYLGTIRRLRRPYRSKYT 128


RESULT 34
AD2441
endopeptidase Clp ATP-binding chain B [imported] - Nostoc sp. (strain PCC 7120)
C;Species: Nostoc sp. PCC 7120
A;Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C;Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Jul-2004
```

```
C;Accession: AD2441
R;Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi
 Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S
DNA Res. 8, 205-213, 2001
A;Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Ana
A;Reference number: AB1807; MUID:21595285; PMID:11759840
A;Accession: AD2441
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-835 <KUR>
A;Cross-references: UNIPROT:Q8YM56; GB:BA000019; PIDN:BAB76783.1; PID:g17134222; GSPDB:G
A;Experimental source: strain PCC 7120
C;Genetics:
A;Gene: clpB
C;Superfamily: endopeptidase Clp ATP-binding chain

  Query Match             27.6%;  Score 50.5;  DB 2;  Length 835;
  Best Local Similarity   38.2%;  Pred. No. 1.6e+02;
  Matches   13;  Conservative    7;  Mismatches   11;  Indels    3;  Gaps    1;

Qy          1 DGLSQEQLEHRERSLQTLRDIQRML---FPDEKE 31
              |  |:|:||  |: |  |:: || |   :  ||:
Db        399 DAASRERLERLEKELADLKEEQRTLNTQWQSEKD 432


RESULT 35
H69337
conserved hypothetical protein AF0704 - Archaeoglobus fulgidus
C;Species: Archaeoglobus fulgidus
C;Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 09-Jul-2004
C;Accession: H69337
R;Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dodson
.; Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E.F.
 Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.
Nature 390, 364-370, 1997
A;Authors: Utterback, T.; Cotton, M.D.; Spriggs, T.; Artiach, P.; Kaine, B.P.; Sykes, S.
Smith, H.O.; Woese, C.R.; Venter, J.C.
A;Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing archaeo
A;Reference number: A69250; MUID:98049343; PMID:9389475
A;Accession: H69337
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-273 <KLE>
A;Cross-references: UNIPROT:O29554; GB:AE001056; GB:AE000782; NID:g2689379; PIDN:AAB9053
C;Superfamily: Methanococcus jannaschii conserved hypothetical protein MJ1557

  Query Match             27.3%;  Score 50;  DB 2;  Length 273;
  Best Local Similarity   37.9%;  Pred. No. 56;
  Matches   11;  Conservative    7;  Mismatches   11;  Indels    0;  Gaps    0;

Qy          1 DGLSQEQLEHRERSLQTLRDIQRMLFPDE 29
              | ||:|:|:   || : : |:  :  |||
Db          9 DRLSEEELKLVRRSFEIIGDVVIIEIPDE 37


RESULT 36
H69843
hypothetical protein yjbH - Bacillus subtilis
C;Species: Bacillus subtilis
C;Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 09-Jul-2004
C;Accession: H69843
R;Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berter
C.; Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Cho
A.; Ehrlich, S.D.; Emmerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.
Nature 390, 249-256, 1997
A;Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Galler
iech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.
Koetter, P.; Koningstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois,
A;Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maueel
y, M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle
 Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scanlon,
A;Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Seror
akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama,
T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K
A;Authors: Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Danchin, A.
A;Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.
A;Reference number: A69580; MUID:98044033; PMID:9384377
A;Accession: H69843
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-275 <KUN>
A;Cross-references: UNIPROT:O31606; GB:Z99110; GB:AL009126; NID:g2633472; PIDN:CAB13012.
A;Experimental source: strain 168
C;Genetics:
A;Gene: yjbH

  Query Match             27.3%;  Score 50;  DB 2;  Length 275;
  Best Local Similarity   37.0%;  Pred. No. 57;
  Matches   10;  Conservative    7;  Mismatches   10;  Indels    0;  Gaps    0;

Qy          7 QLEHRERSLQTLRDIQRMLFPDEKEFT 33
              :|: |:  :| ||::|  ||  :|  |
Db         91 ELQGRKAGMQFLRNMQESLFVSKKNIT 117


RESULT 37
S49771
hypothetical protein YDR175c - yeast (Saccharomyces cerevisiae)
N;Alternate names: hypothetical protein YD9395.08c
C;Species: Saccharomyces cerevisiae
C;Date: 13-Jan-1995 #sequence_revision 10-Feb-1995 #text_change 09-Jul-2004
C;Accession: S49771
R;Murphy, L.; Harris, D.E.
submitted to the EMBL Data Library, November 1994
A;Reference number: S49764
A;Accession: S49771
A;Molecule type: DNA
A;Residues: 1-319 <MUR>
A;Cross-references: UNIPROT:Q03976; EMBL:Z46727; NID:g1289283; PID:e223724; PID:g1289290
C;Genetics:
A;Gene: SGD:RSM24; MIPS:YDR175c
A;Cross-references: SGD:S0002582
A;Map position: 4R

  Query Match             27.3%;  Score 50;  DB 2;  Length 319;
  Best Local Similarity   29.0%;  Pred. No. 67;
  Matches    9;  Conservative    9;  Mismatches   13;  Indels    0;  Gaps    0;

Qy          3 LSQEQLEHRERSLQTLRDIQRMLFPDEKEFT 33
              :| :: ||  :: : | || : |  : |: |
Db        236 MSSDKFEHASQNARYLHDILQRLLAESKDLT 266


RESULT 38
S67595
hypothetical protein YDL060w - yeast (Saccharomyces cerevisiae)
N;Alternate names: hypothetical protein D2544
C;Species: Saccharomyces cerevisiae
C;Date: 12-Jul-1996 #sequence_revision 12-Jul-1996 #text_change 09-Jul-2004
C;Accession: S67595
R;Bloecker, H.; Brandt, P.
submitted to the Protein Sequence Database, July 1996
A;Reference number: S67587
A;Accession: S67595
A;Molecule type: DNA
A;Residues: 1-788 <BLO>
A;Cross-references: UNIPROT:Q07381; EMBL:Z74108; NID:g1431062; PID:g1431063; GSPDB:GN000
A;Experimental source: strain S288C
C;Genetics:
A;Gene: SGD:TSR1; MIPS:YDL060w
A;Cross-references: SGD:S0002218
A;Map position: 4L

  Query Match             27.3%;  Score 50;  DB 2;  Length 788;
  Best Local Similarity   36.4%;  Pred. No. 1.8e+02;
  Matches   12;  Conservative    6;  Mismatches   13;  Indels    2;  Gaps    1;
```

```
Qy          1 DGLSQEQLEHRERSLQTLRDIQR--MLFPDEKE 31
              :|  :   |  || |:  ||:::    |||| |
Db        455 EGFEELSPEEEERQLREFRDMEKEDREFPDEIE 487
```

RESULT 39
D83454
conserved hypothetical protein PA1527 [imported] - Pseudomonas aeruginosa (strain PAO1)
C;Species: Pseudomonas aeruginosa
C;Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 09-Jul-2004
C;Accession: D83454
R;Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrener, P.; Hickey, M.J.; Br
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim,
.; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A;Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pathc
A;Reference number: A82950; MUID:20437337; PMID:10984043
A;Accession: D83454
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1162 <STO>
A;Cross-references: UNIPROT:Q9I3I6; GB:AE004581; GB:AE004091; NID:g9947482; PIDN:AAG0491
A;Experimental source: strain PAO1
C;Genetics:
A;Gene: PA1527

  Query Match   27.3%; Score 50; DB 2; Length 1162;
  Best Local Similarity   39.1%; Pred. No. 2.7e+02;
  Matches   9; Conservative   8; Mismatches   6; Indels   0; Gaps   0;

```
Qy          3 LSQEQLEHRERSLQTLRDIQRML 25
              : | :::| |:||: |:| :| |
Db        400 VQQSRIQHLEQSLERLQDRERRL 422
```

RESULT 40
B36329
hypothetical protein 2 - cabbage looper transposon TED (fragment)
C;Species: Trichoplusia ni (cabbage looper)
C;Date: 01-Feb-1991 #sequence_revision 01-Feb-1991 #text_change 30-Sep-1993
C;Accession: B36329
R;Friesen, P.D.; Nissen, M.S.
Mol. Cell. Biol. 10, 3067-3077, 1990
A;Title: Gene organization and transcription of TED, a lepidopteran retrotransposon inte
A;Reference number: A36329; MUID:90258898; PMID:1692964
A;Accession: B36329
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1236 <FRI>
A;Cross-references: GB:M32662

  Query Match   27.3%; Score 50; DB 2; Length 1236;
  Best Local Similarity   42.9%; Pred. No. 2.9e+02;
  Matches   12; Conservative   4; Mismatches   12; Indels   0; Gaps   0;

```
Qy          5 QEQLEHRERSLQTLRDIQRMLFPDEKEF 32
              || ||: ||  | ||:    :  |: ||
Db        459 QEHLENLERVFQRLRESNFKIQMDKSEF 486
```

Search completed: June 8, 2005, 03:23:55
Job time : 29.5625 secs

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: June 8, 2005, 03:09:59 ; Search time 21.4375 Seconds
(without alignments)
125.671 Million cell updates/sec

Title: US-09-915-543-15_COPY_177_204
Perfect score: 136
Sequence: 1 VYVFSTEMANKAAEAVLKGQVETIVSFH 28

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_79:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	60	44.1	205	2	I40812	porphobilinogen sy
2	55	40.4	634	2	T27465	hypothetical prote
3	51	37.5	243	2	G97070	Zn-dependent hydro
4	49	36.0	210	2	D86398	protein F17L21.2 [
5	49	36.0	458	2	F71315	probable response
6	49	36.0	586	2	D82484	SgaT protein VCA02
7	48.5	35.7	311	2	C81380	probable D-2-hydro
8	48	35.3	330	2	E84074	dihydroxyacetone k
9	47	34.6	319	2	T01822	hypothetical prote
10	47	34.6	352	2	F90179	prolidase (Xaa-Pro
11	47	34.6	662	2	AD0623	probable bacteriop
12	46.5	34.2	586	2	T29695	hypothetical prote
13	46	33.8	330	2	S08500	QUTG protein - Eme
14	46	33.8	363	2	C69962	branched-chain fat
15	46	33.8	461	2	E95887	probable aminotran
16	46	33.8	1289	2	F72308	hypothetical prote
17	46	33.8	2123	2	S55089	probable acetyl-Co
18	45.5	33.5	256	2	S04363	class II histocomp
19	45.5	33.5	268	2	A99261	glutaconate CoA-tr
20	45.5	33.5	504	2	S54744	cellulase (EC 3.2.
21	45.5	33.5	505	2	S39962	endoglucanase - Er
22	45	33.1	131	2	H72478	hypothetical prote
23	45	33.1	451	2	B96495	hypothetical prote
24	45	33.1	555	2	H96762	hypothetical prote
25	44.5	32.7	48	2	D90907	hypothetical prote
26	44.5	32.7	50	2	F85710	unknown protein en
27	44	32.4	126	2	T43131	hypothetical prote
28	44	32.4	265	2	T14645	hypothetical prote
29	44	32.4	318	2	C64445	conserved hypothet
30	44	32.4	340	2	T19105	phosphate carrier
31	44	32.4	350	2	A85056	probable transposo
32	44	32.4	395	2	AH3455	acriflavin resista
33	44	32.4	437	2	A72498	probable DNA/panto
34	44	32.4	476	2	AG1051	probable transport
35	44	32.4	484	2	D65230	hypothetical 52.9
36	44	32.4	484	2	A98275	hypothetical prote
37	44	32.4	484	2	A86116	hypothetical prote
38	44	32.4	533	2	T05092	probable 1,2-diacy
39	44	32.4	609	2	JC5756	vibriolysin (EC 3.
40	44	32.4	619	2	G72709	probable DNA ligas
41	44	32.4	980	2	T39630	valine-tRNA ligase
42	44	32.4	4427	2	PN0637	polyketide synthas
43	43	31.6	119	2	T18644	hypothetical prote
44	43	31.6	159	2	T40440	6,7-dimethyl-8-rib
45	43	31.6	234	2	H98154	amino acid ABC tra

ALIGNMENTS

RESULT 1
I40812
porphobilinogen synthase (EC 4.2.1.24) - Clostridium josui (fragment)
N;Alternate names: delta-aminolevulinic acid dehydratase
C;Species: Clostridium josui
C;Date: 16-Aug-1996 #sequence_revision 16-Aug-1996 #text_change 09-Jul-2004
C;Accession: I40812
R;Fujino, E.; Fujino, T.; Karita, S.; Sakka, K.; Ohmiya, K.
J. Bacteriol. 177, 5169-5175, 1995
A;Title: Cloning and sequencing of some genes responsible for porphyrin biosynthesis fro
A;Reference number: A57344; MUID:95394829; PMID:7665501
A;Accession: I40812
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-205 <RES>
A;Cross-references: UNIPROT:Q59295; GB:D28503; NID:g536874; PIDN:BAA05863.1; PID:g556484
C;Genetics:
A;Gene: hemB
C;Superfamily: porphobilinogen synthase
C;Keywords: carbon-oxygen lyase; hydro-lyase

Query Match 44.1%; Score 60; DB 2; Length 205;
Best Local Similarity 46.2%; Pred. No. 0.24;
Matches 12; Conservative 5; Mismatches 9; Indels 0; Gaps 0;

```
Qy          2 YVFSTEMANKAAEAVLKGQVETIVSF 27
              | || :|  || || ||  |:::: |
Db         51 YHFSPDMVGKAIEAALKADVKSVLLF 76
```

RESULT 2
T27465
hypothetical protein Y87G2A.m - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C;Accession: T27465
R;White, S.
submitted to the EMBL Data Library, September 1999
A;Reference number: Z20371
A;Accession: T27465
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-634 <WIL>
A;Cross-references: EMBL:AL110500; NID:e1542314; PIDN:CAB54487.1; CESP:Y87G2A.m
A;Experimental source: clone Y87G2A
C;Genetics:
A;Gene: CESP:Y87G2A.m
A;Introns: 74/1; 270/1

Query Match 40.4%; Score 55; DB 2; Length 634;
Best Local Similarity 44.4%; Pred. No. 4.4;
Matches 12; Conservative 2; Mismatches 13; Indels 0; Gaps 0;

```
Qy          2 YVFSTEMANKAAEAVLKGQVETIVSFH 28
              ||    || ||  ||  | :: |  ||
Db        461 YVKCAHMAEKAVAAVANGDLQIIPEFH 487
```

RESULT 3
G97070
Zn-dependent hydrolases, glyoxylase family [imported] - Clostridium acetobutylicum
C;Species: Clostridium acetobutylicum
C;Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 09-Jul-2004
C;Accession: G97070
R;Nolling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee,
.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.
J. Bacteriol. 183, 4823-4838, 2001
A;Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Clo
A;Reference number: A96900; MUID:21359325; PMID:21359325
A;Accession: G97070
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-243 <KUR>
A;Cross-references: UNIPROT:Q97JA0; GB:AE001437; PIDN:AAK79354.1; PID:g15024323; GSPDB:G
A;Experimental source: Clostridium acetobutylicum ATCC824
C;Genetics:
A;Gene: CAC1386

```
  Query Match             37.5%;  Score 51;  DB 2;  Length 243;
  Best Local Similarity   25.0%;  Pred. No. 6.2;
  Matches    7;  Conservative   11;  Mismatches   10;  Indels    0;  Gaps    0;

Qy          1 VYVFSTEMANKAAEAVLKGQVETIVSFH 28
              :: | : :: |: | : |  :||:: :|
Db        199 LFDFDSNLSKKSLEKLTKYDIETVICYH 226
```

RESULT 4
D86398
protein F17L21.2 [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 09-Jul-2004
C;Accession: D86398
R;Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
 Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;
ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A;Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziali,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A;Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A;Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A;Reference number: A86141; MUID:21016719; PMID:11130712
A;Accession: D86398
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-210 <STO>
A;Cross-references: UNIPROT:Q9FZL1; GB:AE005172; NID:g9802520; PIDN:AAF99722.1; GSPDB:GN
C;Genetics:
A;Gene: F17L21.2
A;Map position: 1

```
  Query Match             36.0%;  Score 49;  DB 2;  Length 210;
  Best Local Similarity   41.7%;  Pred. No. 11;
  Matches   10;  Conservative    6;  Mismatches    8;  Indels    0;  Gaps    0;

Qy          4 FSTEMANKAAEAVLKGQVETIVSF 27
              :|    ||||: : ||::|  |:|
Db        181 WSFRSTNKAADRLAKGELENNVTF 204
```

RESULT 5
F71315
probable response regulatory protein (atoC) - syphilis spirochete
C;Species: Treponema pallidum subsp. pallidum (syphilis spirochete)
C;Date: 24-Jul-1998 #sequence_revision 24-Jul-1998 #text_change 02-Jun-2003
C;Accession: F71315
R;Fraser, C.M.; Norris, S.J.; Weinstock, G.M.; White, O.; Sutton, G.G.; Dodson, R.; Gwinn
rson, J.; Khalak, H.; Richardson, D.; Howell, J.K.; Chidambaram, M.; Utterback, T.; McDon
they, L.; Weidman, J.; Smith, H.O.; Venter, J.C.
Science 281, 375-388, 1998
A;Title: Complete genome sequence of Treponema pallidum, the syphilis spirochete.
A;Reference number: A71250; MUID:98332770; PMID:9665876
A;Accession: F71315
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-458 <COL>
A;Cross-references: GB:AE001227; GB:AE000520; NID:g3322797; PIDN:AAC65507.1; PID:g3322811
A;Experimental source: strain Nichols
C;Genetics:
A;Gene: TP0519
C;Superfamily: response regulator, NtrC type; response regulator homology; RNA polymerase
C;Keywords: phosphoprotein
F;5-114/Domain: response regulator homology <RRH>
F;143-365/Domain: RNA polymerase sigma factor interaction domain homology <SFI>
F;53/Binding site: phosphate (Asp) (covalent) #status predicted
F;74/Binding site: phosphate (Asp) (covalent) #status predicted

```
  Query Match             36.0%;  Score 49;  DB 2;  Length 458;
  Best Local Similarity   37.5%;  Pred. No. 24;
  Matches    9;  Conservative    5;  Mismatches   10;  Indels    0;  Gaps    0;

Qy          3 VFSTEMANKAAEAVLKGQVETIVS 26
              ||: |  |   |  ||| :: |::
Db         29 VFTAEDGNTGVEIALKGDIDLIIT 52
```

RESULT 6
D82484
SgaT protein VCA0246 [imported] - Vibrio cholerae (strain N16961 serogroup O1)
C;Species: Vibrio cholerae
C;Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 09-Jul-2004
C;Accession: D82484
R;Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.J.;
chardson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragoi, I.; Sellers, P.
l, R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.
Nature 406, 477-483, 2000
A;Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.
A;Reference number: A82035; MUID:20406833; PMID:10952301
A;Accession: D82484
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-586 <HEI>
A;Cross-references: UNIPROT:Q9KMS4; GB:AE004364; GB:AE003853; NID:g9657630; PIDN:AAF9615
A;Experimental source: serogroup O1; strain N16961; biotype El Tor
C;Genetics:
A;Gene: VCA0246
A;Map position: 2

```
  Query Match             36.0%;  Score 49;  DB 2;  Length 586;
  Best Local Similarity   25.0%;  Pred. No. 31;
  Matches   11;  Conservative    8;  Mismatches    7;  Indels   18;  Gaps    1;

Qy          2 YVFSTEMANKA----------------AEAVLKGQVETIVSF 27
              |:| :::  ||                |  ::|| ::||| |
Db          9 YIFYSQVMTKAPLLLGLVTLIGYWLLRRDATTIIKGSIKTIVGF 52
```

RESULT 7
C81380
probable D-2-hydroxyacid dehydrogenase Cj0373 [imported] - Campylobacter jejuni (strain N
C;Species: Campylobacter jejuni
C;Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 09-Jul-2004
C;Accession: C81380
R;Parkhill, J.; Wren, B.W.; Mungall, K.; Ketley, J.M.; Churcher, C.; Basham, D.; Chilling
C.W.; Quail, M.; Rajandream, M.A.; Rutherford, K.M.; VanVliet, A.; Whitehead, S.; Barrell
Nature 403, 665-668, 2000

```
A;Title: The genome sequence of the food-borne pathogen Campylobacter jejuni reveals hyp
A;Reference number: A81250; MUID:20150912; PMID:10688204
A;Accession: C81380
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-311 <PAR>
A;Cross-references: UNIPROT:Q9PIC9; GB:AL139075; GB:AL111168; NID:g6967817; PIDN:CAB7420
A;Experimental source: serotype O2, strain NCTC 11168
C;Genetics:
A;Gene: Cj0373
C;Superfamily: phosphoglycerate dehydrogenase

  Query Match             35.7%;  Score 48.5;  DB 2;  Length 311;
  Best Local Similarity   42.4%;  Pred. No. 19;
  Matches   14;  Conservative    6;  Mismatches    8;  Indels    5;  Gaps    2;

Qy          1 VYVFSTEMANKAAEAV---LKGQVET--IVSFH 28
              :| :||  ||| |: |   ||  ::|  |:| |
Db        171 IYYYSTSGANKNADFVHLELKDLLKTCDIISIH 203


RESULT 8
E84074
dihydroxyacetone kinase BH3397 [imported] - Bacillus halodurans (strain C-125)
C;Species: Bacillus halodurans
C;Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 09-Jul-2004
C;Accession: E84074
R;Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fuji, F.; Hira
Nucleic Acids Res. 28, 4317-4331, 2000
A;Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and
A;Reference number: A83650; MUID:20512582; PMID:11058132
A;Accession: E84074
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-330 <STO>
A;Cross-references: UNIPROT:Q9K7G4; GB:AP001518; GB:BA000004; NID:g10175792; PIDN:BAB071
A;Experimental source: strain C-125
C;Genetics:
A;Gene: BH3397

  Query Match             35.3%;  Score 48;  DB 2;  Length 330;
  Best Local Similarity   42.3%;  Pred. No. 24;
  Matches   11;  Conservative    4;  Mismatches   11;  Indels    0;  Gaps    0;

Qy          2 YVFSTEMANKAAEAVLKGQVETIVSF 27
              |||  ::|||  |  |  | : : ||
Db        270 YVFMNDVANKLTEEGLNIQFKKVGSF 295


RESULT 9
T01822
hypothetical protein T27D20.16 - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 26-Feb-1999 #sequence_revision 26-Feb-1999 #text_change 09-Jul-2004
C;Accession: T01822
R;Edwards, J.; Wollam, C.; Dubbelde, C.
submitted to the EMBL Data Library, August 1998
A;Description: The sequence of A. thaliana T27D20.
A;Reference number: Z14441
A;Accession: T01822
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-319 <EDW>
A;Cross-references: UNIPROT:O81460; EMBL:AF076274; NID:g3293583; PID:g3377852
A;Experimental source: cultivar Columbia
C;Genetics:
A;Map position: 4
A;Introns: 49/3; 151/3; 210/3; 269/2
A;Note: T27D20.16
C;Superfamily: Arabidopsis hypothetical protein F7N22.18

  Query Match             34.6%;  Score 47;  DB 2;  Length 319;
  Best Local Similarity   48.0%;  Pred. No. 33;
  Matches   12;  Conservative    4;  Mismatches    7;  Indels    2;  Gaps    1;

Qy          5 STEMANK--AAEAVLKGQVETIVSF 27
              | |:: |  ||||::  | | | ||
Db        243 SIELSQKLAAAEALIANQAEKITSF 267


RESULT 10
F90179
prolidase (Xaa-Pro dipeptidase) (pepQ) [imported] - Sulfolobus solfataricus
C;Species: Sulfolobus solfataricus
C;Date: 24-May-2001 #sequence_revision 24-May-2001 #text_change 09-Jul-2004
C;Accession: F90179
R;She, Q.; Singh, R.K.; Confalonieri, F.; Zivanovic, Y.; Allard, G.; Awayez, M.J.; Chan-
 Jong, I.; Jeffries, A.C.; Kozera, C.J.; Medina, N.; Peng, X.; Thi-Ngoc, H.P.; Redder, P
arrett, R.A.; Ragan, M.A.; Sensen, C.W.; Van der Oost, J.
submitted to GenBank, April 2001
A;Description: Sulfolobus solfataricus complete genome.
A;Reference number: A99139
A;Accession: F90179
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-352 <KUR>
A;Cross-references: UNIPROT:Q980E9; GB:AE006641; NID:g13813507; PIDN:AAK40693.1; GSPDB:G
C;Genetics:
A;Gene: pepQ
C;Superfamily: X-Pro aminopeptidase

  Query Match             34.6%;  Score 47;  DB 2;  Length 352;
  Best Local Similarity   43.5%;  Pred. No. 36;
  Matches   10;  Conservative    3;  Mismatches   10;  Indels    0;  Gaps    0;

Qy          2 YVFSTEMANKAAEAVLKGQVETI 24
              :||    | |  | ||: |:| |
Db        232 FVFKNSEAKKVYEVVLEAQMEAI 254


RESULT 11
AD0623
probable bacteriophage protein STY1061 [imported] - Salmonella enterica subsp. enterica
C;Species: Salmonella enterica subsp. enterica serovar Typhi
A;Note: this species has also been called Salmonella typhi
C;Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 18-Nov-2002
C;Accession: AD0623
R;Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher,
th, T.; Connerton, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar,
, S.; Moule, S.; O'Gaora, P.
Nature 413, 848-852, 2001
A;Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.;
A;Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serov
A;Reference number: AB0502; MUID:21534947; PMID:11677608
A;Accession: AD0623
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-662 <PAR>
A;Cross-references: GB:AL513382; PIDN:CAD05454.1; PID:g16502215; GSPDB:GN00176
C;Genetics:
A;Gene: STY1061

  Query Match             34.6%;  Score 47;  DB 2;  Length 662;
  Best Local Similarity   40.9%;  Pred. No. 71;
  Matches    9;  Conservative    6;  Mismatches    7;  Indels    0;  Gaps    0;

Qy          3 VFSTEMANKAAEAVLKGQVETI 24
              ::| |: |||| | : |:|  :
Db        600 IYSRELINKAAVAGISGKVTEV 621


RESULT 12
T29695
hypothetical protein T18H9.1 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
```

```
C;Accession: T29695
R;Du, Z.; Gattung, S.
submitted to the EMBL Data Library, November 1995
A;Description: The sequence of C. elegans cosmid T18H9.
A;Reference number: Z20666
A;Accession: T29695
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-586 <DUZ>
A;Cross-references: UNIPROT:Q22551; EMBL:U41746; PIDN:AAA83334.1; CESP:T18H9.1
C;Genetics:
A;Gene: CESP:T18H9.1
A;Introns: 23/1; 105/3; 174/3; 303/3; 403/2; 430/3; 539/2

  Query Match             34.2%;  Score 46.5;  DB 2;  Length 586;
  Best Local Similarity   40.7%;  Pred. No. 74;
  Matches   11;  Conservative    7;  Mismatches    8;  Indels    1;  Gaps    1;

Qy          1 VYVFSTEMANKAAEAVLKGQVETIVSF 27
              |: |:|::  | :| :   | |||||:
Db        492 VHKFATKL-QKHSEKIFSTQFETIVSY 517


RESULT 13
S08500
QUTG protein - Emericella nidulans
C;Species: Emericella nidulans, Aspergillus nidulans
C;Date: 21-Nov-1993 #sequence_revision 01-Dec-1995 #text_change 09-Jul-2004
C;Accession: S08500; S11944
R;Hawkins, A.R.; Lamb, H.K.; Smith, M.; Keyte, J.W.; Roberts, C.F.
Mol. Gen. Genet. 214, 224-231, 1988
A;Title: Molecular organisation of the quinic acid utilization (QUT) gene cluster in Asp
A;Reference number: S08498; MUID:89181521; PMID:2976880
A;Accession: S08500
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-330 <HAW>
A;Cross-references: UNIPROT:P25416; EMBL:X13525; NID:g2398; PID:g2399
A;Note: the authors translated the codon GCG for residue 327 as Thr
R;Lamb, H.K.; Hawkins, A.R.; Smith, M.; Harvey, I.J.; Brown, J.; Turner, G.; Roberts, C.
Mol. Gen. Genet. 223, 17-23, 1990
A;Title: Spatial and biological characterisation of the complete quinic acid utilisation
A;Reference number: S11944; MUID:91080861; PMID:2175387
A;Accession: S11944
A;Molecule type: DNA
A;Residues: 1-326,'T' <LAM>
A;Experimental source: strain 2035
C;Genetics:
A;Introns: 57/1; 104/1; 250/2; 293/2

  Query Match             33.8%;  Score 46;  DB 2;  Length 330;
  Best Local Similarity   38.9%;  Pred. No. 47;
  Matches    7;  Conservative    7;  Mismatches    4;  Indels    0;  Gaps    0;

Qy          1 VYVFSTEMANKAAEAVLK 18
              :| |:|::| || : :|:
Db         13 IYAFATDLARKAGQLLLE 30


RESULT 14
C69962
branched-chain fatty-acid kinase homolog yqiU - Bacillus subtilis
C;Species: Bacillus subtilis
C;Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 09-Jul-2004
C;Accession: C69962
R;Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berter
C.; Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Cho
A.; Ehrlich, S.D.; Emmerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.
Nature 390, 249-256, 1997
A;Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Galler
iech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.
Koetter, P.; Koningstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois,
A;Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maueel
y, M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle,
 Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scanlon,
A;Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Seror,
akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama,
T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K
A;Authors: Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Danchin, A.
A;Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.
A;Reference number: A69580; MUID:98044033; PMID:9384377
A;Accession: C69962
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-363 <KUN>
A;Cross-references: UNIPROT:P54532; GB:Z99116; GB:AL009126; NID:g2634723; PIDN:CAB14338.1
A;Experimental source: strain 168
C;Genetics:
A;Gene: yqiU
C;Superfamily: butyrate kinase

  Query Match             33.8%;  Score 46;  DB 2;  Length 363;
  Best Local Similarity   71.4%;  Pred. No. 52;
  Matches   10;  Conservative    1;  Mismatches    3;  Indels    0;  Gaps    0;

Qy         12 AAEAVLKGQVETIV 25
              || | |||:|| ||
Db        293 AASAALKGEVEAIV 306


RESULT 15
E95887
probable aminotransferase protein [imported] - Sinorhizobium meliloti (strain 1021) magap
C;Species: Sinorhizobium meliloti
C;Date: 24-Aug-2001 #sequence_revision 24-Aug-2001 #text_change 09-Jul-2004
C;Accession: E95887
R;Finan, T.M.; Weidner, S.; Wong, K.; Buhrmester, J.; Chain, P.; Vorholter, F.J.; Hernanc
Proc. Natl. Acad. Sci. U.S.A. 98, 9889-9894, 2001
A;Title: The complete sequence of the 1,683-kb pSymB megaplasmid from the N2-fixing endos
A;Reference number: A95842; MUID:21396508; PMID:11481431
A;Accession: E95887
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-461 <KUR>
A;Cross-references: UNIPROT:Q92WH4; GB:AL591985; PIDN:CAC48765.1; PID:g15140238; GSPDB:GN
A;Experimental source: strain 1021, megaplasmid pSymB
R;Galibert, F.; Finan, T.M.; Long, S.R.; Puhler, A.; Abola, P.; Ampe, F.; Barloy-Hubler,
pela, D.; Chain, P.; Cowie, A.; Davis, R.W.; Dreano, S.; Federspiel, N.A.; Fisher, R.F.;
 L.; Hyman, R.W.; Jones, T.
Science 293, 668-672, 2001
A;Authors: Kahn, D.; Kahn, M.L.; Kalman, S.; Keating, D.H.; Kiss, E.; Komp, C.; Lelaure,
hebault, P.; Vandenbol, M.; Vorholter, F.J.; Weidner, S.; Wells, D.H.; Wong, K.; Yeh, K.C
A;Title: The composite genome of the legume symbiont Sinorhizobium meliloti.
A;Reference number: A96039; MUID:21368234; PMID:11474104
A;Contents: annotation
C;Genetics:
A;Gene: SMb20379
A;Genome: plasmid

  Query Match             33.8%;  Score 46;  DB 2;  Length 461;
  Best Local Similarity   41.7%;  Pred. No. 68;
  Matches   10;  Conservative    4;  Mismatches   10;  Indels    0;  Gaps    0;

Qy          4 FSTEMANKAAEAVLKGQVETIVSF 27
              |:| |||     :|:   ||| :|
Db        199 FATRMANNLEAIILEEGPETIAAF 222


RESULT 16
F72308
hypothetical protein - Thermotoga maritima (strain MSB8)
C;Species: Thermotoga maritima
C;Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 09-Jul-2004
C;Accession: F72308
R;Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwinn, M.L.; Dodson, R.J.; Haft, D.H.; Hickey,
Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson, D.;
```

C.M.
Nature 399, 323-329, 1999
A;Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome seq
A;Reference number: A72200; MUID:99287316; PMID:10360571
A;Accession: F72308
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1289 <ARN>
A;Cross-references: UNIPROT:Q9X087; GB:AE001761; GB:AE000512; NID:g4981529; PIDN:AAD3607
A;Experimental source: strain MSB8
C;Genetics:
A;Gene: TM0992

Query Match 33.8%; Score 46; DB 2; Length 1289;
Best Local Similarity 52.9%; Pred. No. 2e+02;
Matches 9; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

Qy 7 EMANKAAEAVLKGQVET 23
Db 264 ETAKKSAESILKNIIET 280

RESULT 17
S55089
probable acetyl-CoA carboxylase (EC 6.4.1.2) HFA1 - yeast (Saccharomyces cerevisiae)
N;Alternate names: protein YM8261.01c; protein YM8325.08c; protein YMR207c
C;Species: Saccharomyces cerevisiae
C;Date: 08-Jul-1995 #sequence_revision 05-Dec-1997 #text_change 09-Jul-2004
C;Accession: S55089; S41802; S59447
R;Dedman, K.; Brown, D.; Bowman, S.
submitted to the EMBL Data Library, June 1995
A;Reference number: S55089
A;Accession: S55089
A;Molecule type: DNA
A;Residues: 1-833 <DED>
A;Cross-references: UNIPROT:P32874; EMBL:Z49809; MIPS:YMR207c; NID:g854459; PIDN:CAA8992
A;Experimental source: strain AB972
A;Note: the published sequence extends beyond the amino end
R;Kearsey, S.E.
submitted to the EMBL Data Library, April 1993
A;Description: Identification of an Saccharomyces cerevisiae gene closely related to FAS
A;Reference number: S41802
A;Accession: S41802
A;Molecule type: DNA
A;Residues: 1-510,'L',512-799 <KEA>
A;Cross-references: EMBL:Z22558; NID:g296212; PIDN:CAA80280.1; PID:g388250
A;Note: the published sequence extends beyond the amino end
R;Odell, C.; Bowman, S.
submitted to the EMBL Data Library, March 1995
A;Reference number: S59441
A;Accession: S59447
A;Molecule type: DNA
A;Residues: 812-2123 <ODE>
A;Cross-references: EMBL:Z48755; MIPS:YMR207c; NID:g736296; PIDN:CAA88647.1; PID:g763183
A;Experimental source: strain AB972
C;Genetics:
A;Gene: SGD:HFA1
A;Cross-references: MIPS:YMR207c; SGD:S0004820
A;Map position: 13R
C;Superfamily: human acetyl-CoA carboxylase; biotin carboxylase homology; lipoyl/biotin-
C;Keywords: biotin binding; ligase
F;1-487/Domain: biotin carboxylase homology <BCH>
F;615-687/Domain: lipoyl/biotin-binding homology <LPB>
F;654/Binding site: biotin (Lys) (covalent) #status predicted

Query Match 33.8%; Score 46; DB 2; Length 2123;
Best Local Similarity 45.0%; Pred. No. 3.4e+02;
Matches 9; Conservative 5; Mismatches 6; Indels 0; Gaps 0;

Qy 2 YVFSTEMANKAAEAVLKGQV 21
Db 509 YVFTEKVRNKYLELLRRGQV 528

RESULT 18
S04363
class II histocompatibility antigen RT1-B alpha chain precursor - rat
C;Species: Rattus norvegicus (Norway rat)
C;Date: 28-Feb-1990 #sequence_revision 28-Feb-1990 #text_change 09-Jul-2004
C;Accession: S04363
R;Syha, J.; Henkes, W.; Reske, K.
Nucleic Acids Res. 17, 3985, 1989
A;Title: Complete cDNA sequence coding for the MHC class II RT1.B alpha chain of the LEW
A;Reference number: S04363; MUID:89282410; PMID:2499874
A;Accession: S04363
A;Molecule type: mRNA
A;Residues: 1-256 <SYH>
A;Cross-references: UNIPROT:Q95572; EMBL:X14879; NID:g57154; PIDN:CAA33020.1; PID:g57155
C;Superfamily: class II histocompatibility antigen; immunoglobulin homology
F;1-23/Domain: signal sequence #status predicted <SIG>
F;24-256/Product: class II histocompatibility antigen, RT1-B alpha chain #status predict
F;127-192/Domain: immunoglobulin homology <IMM>

Query Match 33.5%; Score 45.5; DB 2; Length 256;
Best Local Similarity 40.0%; Pred. No. 43;
Matches 12; Conservative 3; Mismatches 8; Indels 7; Gaps 1;

Qy 5 STEMANKAAEA-------VLKGQVETIVSF 27
Db 106 STQAVNKVPEATVFSKSPVLLGQPNTLICF 135

RESULT 19
A99261
glutaconate CoA-transferase, subunit A (gctA) [imported] - Sulfolobus solfataricus
C;Species: Sulfolobus solfataricus
C;Date: 24-May-2001 #sequence_revision 24-May-2001 #text_change 09-Jul-2004
C;Accession: A99261
R;She, Q.; Singh, R.K.; Confalonieri, F.; Zivanovic, Y.; Allard, G.; Awayez, M.J.; Chan-
Jong, I.; Jeffries, A.C.; Kozera, C.J.; Medina, N.; Peng, X.; Thi-Ngoc, H.P.; Redder, P
arrett, R.A.; Ragan, M.A.; Sensen, C.W.; Van der Oost, J.
submitted to GenBank, April 2001
A;Description: Sulfolobus solfataricus complete genome.
A;Reference number: A99139
A;Accession: A99261
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-268 <KUR>
A;Cross-references: UNIPROT:Q97Z52; GB:AE006641; NID:g13814274; PIDN:AAK41344.1; GSPDB:G
C;Genetics:
A;Gene: gctA

Query Match 33.5%; Score 45.5; DB 2; Length 268;
Best Local Similarity 33.3%; Pred. No. 45;
Matches 13; Conservative 5; Mismatches 10; Indels 11; Gaps 2;

Qy 1 VYVFSTEMAN---------KAAEAVLKGQVETI--VSFH 28
Db 6 IYIFFVEIINMESKLLSLEEAVEVVKKGDSVTISGISIH 44

RESULT 20
S54744
cellulase (EC 3.2.1.4) CelV1 precursor - Erwinia carotovora (SCC 3193)
N;Alternate names: endo-1,4-beta-glucanase
C;Species: Erwinia carotovora
A;Variety: SCC 3193
C;Date: 27-Oct-1995 #sequence_revision 03-Nov-1995 #text_change 09-Jul-2004
C;Accession: S54744; S44996
R;Maee, A.; Heikinheimo, R.; Palva, E.T.
Mol. Gen. Genet. 247, 17-26, 1995
A;Title: Structure and regulation of the Erwinia carotovora subspecies carotovora SCC319
A;Reference number: S54744; MUID:95231512; PMID:7715600
A;Accession: S54744
A;Molecule type: DNA
A;Residues: 1-504 <MAE>

```
A;Cross-references: UNIPROT:Q59395; EMBL:X79241; NID:g493492; PIDN:CAA55823.1; PID:g4934
C;Genetics:
A;Gene: celV1
C;Function:
A;Description: hydrolysis of 1,4-beta-D-glucosidic linkages in beta-D-glucans such as ce
A;Pathway: cellulose degradation
C;Keywords: glycosidase; hydrolase; polysaccharide degradation
F;1-32/Domain: signal sequence #status predicted <SIG>
F;33-504/Product: cellulase #status predicted <MAT>

  Query Match             33.5%;  Score 45.5;  DB 2;  Length 504;
  Best Local Similarity   36.7%;  Pred. No. 88;
  Matches   11;  Conservative    5;  Mismatches   11;  Indels    3;  Gaps    1;

Qy          2 YVFSTEMANKAAEAVLKGQ---VETIVSFH 28
              |: :  :|||  |||   |   |  |: :|
Db        101 YIANPSLANKVKEAVAAAQGLGVYIIIDWH 130


RESULT 21
S39962
endoglucanase - Erwinia carotovora
C;Species: Erwinia carotovora
C;Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 09-Jul-2004
C;Accession: S39962
R;Cooper, V.J.C.; Salmond, G.P.C.
Mol. Gen. Genet. 241, 341-350, 1993
A;Title: Molecular analysis of the major cellulase (CelV) of Erwinia carotovora: evidenc
A;Reference number: S39962; MUID:94067016; PMID:8246888
A;Accession: S39962
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-505 <COO>
A;Cross-references: UNIPROT:Q47096; EMBL:X76000; NID:g434941; PIDN:CAA53592.1; PID:g4349

  Query Match             33.5%;  Score 45.5;  DB 2;  Length 505;
  Best Local Similarity   36.7%;  Pred. No. 88;
  Matches   11;  Conservative    5;  Mismatches   11;  Indels    3;  Gaps    1;

Qy          2 YVFSTEMANKAAEAVLKGQ---VETIVSFH 28
              |: :  :|||  |||   |   |  |: :|
Db        101 YISNPSLANKVKEAVAAAQSLGVYIIIDWH 130


RESULT 22
H72478
hypothetical protein APE2472 - Aeropyrum pernix (strain K1)
C;Species: Aeropyrum pernix
C;Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 09-Jul-2004
C;Accession: H72478
R;Kawarabayasi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Haikawa, Y.; Jin-no, K.; Takah
awa, H.; Takamiya, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; K
DNA Res. 6, 83-101, 1999
A;Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, Aeropyr
A;Reference number: A72450; MUID:99310339; PMID:10382966
A;Accession: H72478
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-131 <KAW>
A;Cross-references: UNIPROT:Q9Y912; DDBJ:AP000064; NID:g5105945; PIDN:BAA81488.1; PID:d1
A;Experimental source: strain K1
C;Genetics:
A;Gene: APE2472
C;Superfamily: Aeropyrum pernix hypothetical protein APE2472

  Query Match             33.1%;  Score 45;  DB 2;  Length 131;
  Best Local Similarity   47.6%;  Pred. No. 25;
  Matches   10;  Conservative    4;  Mismatches    7;  Indels    0;  Gaps    0;

Qy          1 VYVFSTEMANKAAEAVLKGQV 21
              ||:|| |:  :| || : | |
Db         62 VYLFSGEVWRRAVEATMIGGV 82


RESULT 23
B96495
hypothetical protein T8D8.9 [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 09-Jul-2004
C;Accession: B96495
R;Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
 Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;
ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A;Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziali,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A;Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon, I
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A;Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A;Reference number: A86141; MUID:21016719; PMID:11130712
A;Accession: B96495
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-451 <STO>
A;Cross-references: UNIPROT:Q9C8E8; GB:AE005173; NID:g11141992; PIDN:AAG32071.1; GSPDB:GN
C;Genetics:
A;Gene: T8D8.9
A;Map position: 1
C;Superfamily: Arabidopsis hypothetical protein F7N22.18

  Query Match             33.1%;  Score 45;  DB 2;  Length 451;
  Best Local Similarity   48.0%;  Pred. No. 93;
  Matches   12;  Conservative    4;  Mismatches    7;  Indels    2;  Gaps    1;

Qy          3 VFSTEMANKAAEAVLKGQVETIVSF 27
              | | ::|  ||||::  | | | ||
Db        361 VLSQKLA--AAEALIANQAEKITSF 383


RESULT 24
H96762
hypothetical protein F6D5.1 [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
C;Accession: H96762
R;Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
 Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;
ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A;Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziali,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A;Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon, I
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A;Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A;Reference number: A86141; MUID:21016719; PMID:11130712
A;Accession: H96762
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-555 <STO>
A;Cross-references: GB:AE005173; NID:g10092368; PIDN:AAG12776.1; GSPDB:GN00141
C;Genetics:
A;Gene: F6D5.1
A;Map position: 1

  Query Match             33.1%;  Score 45;  DB 2;  Length 555;
  Best Local Similarity   31.8%;  Pred. No. 1.2e+02;
  Matches    7;  Conservative    8;  Mismatches    7;  Indels    0;  Gaps    0;

Qy          4 FSTEMANKAAEAVLKGQVETIV 25
              |:||:| || : :    :|:::
Db        131 FTTELAQKAGQVIAVDFIESVI 152


RESULT 25
```

```
D90907
hypothetical protein ECs2228 [imported] - Escherichia coli (strain O157:H7, substrain RI
C;Species: Escherichia coli
C;Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 09-Jul-2004
C;Accession: D90907
R;Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.
gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001
A;Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and genc
A;Reference number: A99629; MUID:21156231; PMID:11258796
A;Accession: D90907
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-48 <HAY>
A;Cross-references: UNIPROT:Q8XC69; GB:BA000007; PIDN:BAB35651.1; PID:g13361694; GSPDB:G
A;Experimental source: strain O157:H7, substrain RIMD 0509952
C;Genetics:
A;Gene: ECs2228

  Query Match             32.7%;  Score 44.5;  DB 2;  Length 48;
  Best Local Similarity   43.5%;  Pred. No. 10;
  Matches   10;  Conservative    5;  Mismatches    5;  Indels    3;  Gaps    1;

Qy          2 YVFSTEMANKAAE---AVLKGQV 21
              |: ||  ||: ||    |::||:
Db          7 YISSTSFANEMAEMRQQVMEGQI 29


RESULT 26
F85710
unknown protein encoded by prophage CP-933O [imported] - Escherichia coli (strain O157:H
C;Species: Escherichia coli
C;Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 09-Jul-2004
C;Accession: F85710
R;Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
iller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,
Nature 409, 529-533, 2001
A;Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A;Reference number: A85480; MUID:21074935; PMID:11206551
A;Accession: F85710
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-50 <STO>
A;Cross-references: UNIPROT:Q8XC69; GB:AE005174; NID:g12515022; PIDN:AAG56146.1; GSPDB:G
A;Experimental source: strain O157:H7, substrain EDL933
C;Genetics:
A;Gene: Z2076

  Query Match             32.7%;  Score 44.5;  DB 2;  Length 50;
  Best Local Similarity   43.5%;  Pred. No. 11;
  Matches   10;  Conservative    5;  Mismatches    5;  Indels    3;  Gaps    1;

Qy          2 YVFSTEMANKAAE---AVLKGQV 21
              |: ||  ||: ||    |::||:
Db          9 YISSTSFANEMAEMRQQVMEGQI 31


RESULT 27
T43131
hypothetical protein - Lactococcus lactis plasmid pMRC01
C;Species: Lactococcus lactis
C;Date: 02-Jun-2000 #sequence_revision 02-Jun-2000 #text_change 09-Jul-2004
C;Accession: T43131
R;Dougherty, B.A.; Hill, C.; Weidman, J.F.; Richardson, D.R.; Venter, J.C.; Ross, R.P.
Mol. Microbiol. 29, 1029-1038, 1998
A;Title: Sequence and analysis of the 60 kb conjugative, bacteriocin-producing plasmid p
A;Reference number: Z22314
A;Accession: T43131
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-126 <DOU>
A;Cross-references: UNIPROT:O32777; EMBL:AE001272; PIDN:AAC56049.1
A;Experimental source: strain DPC3147
C;Genetics:
A;Genome: plasmid pMRC01
A;Note: ORF00060

  Query Match             32.4%;  Score 44;  DB 2;  Length 126;
  Best Local Similarity   33.3%;  Pred. No. 34;
  Matches    7;  Conservative    8;  Mismatches    6;  Indels    0;  Gaps    0;

Qy          8 MANKAAEAVLKGQVETIVSFH 28
              :|:: || | || : ::: :|
Db         67 IASRIAEYVTKGSLVSLIGYH 87


RESULT 28
T14645
hypothetical protein 265 - Sorghum mitochondrion
C;Species: mitochondrion Sorghum bicolor (sorghum)
C;Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 09-Jul-2004
C;Accession: T14645
R;Tang, H.V.; Pring, D.R.; Muza, F.R.; Yan, B.
Curr. Genet. 29, 265-274, 1996
A;Title: Sorghum mitochondrial orf25 and a related chimeric configuration of a male-ster
A;Reference number: S65767; MUID:96163056; PMID:8595673
A;Accession: T14645
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-265 <VAN>
A;Cross-references: UNIPROT:Q35783; EMBL:U22068; NID:g733079; PIDN:AAA97555.1; PID:g7330
A;Experimental source: strain IS1112C; coleoptile
C;Genetics:
A;Genome: mitochondrion
C;Keywords: mitochondrion

  Query Match             32.4%;  Score 44;  DB 2;  Length 265;
  Best Local Similarity   37.5%;  Pred. No. 74;
  Matches    9;  Conservative    6;  Mismatches    9;  Indels    0;  Gaps    0;

Qy          1 VYVFSTEMANKAAEAVLKGQVETI 24
              |:||| :   |  :  | |::|:|
Db         22 VWVFSRKSLGKTFKETLDGRIESI 45


RESULT 29
C64445
conserved hypothetical protein MJ1164 - Methanococcus jannaschii
C;Species: Methanococcus jannaschii
C;Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 09-Jul-2004
C;Accession: C64445
R;Bult, C.J.; White, O.; Olsen, G.J.; Zhou, L.; Fleischmann, R.D.; Sutton, G.G.; Blake,
; Reich, C.I.; Overbeek, R.; Kirkness, E.F.; Weinstock, K.G.; Merrick, J.M.; Glodek, A.;
rson, J.D.; Sadow, P.W.; Hanna, M.C.; Cotton, M.D.; Roberts, K.M.; Hurst, M.A.
Science 273, 1058-1073, 1996
A;Authors: Kaine, B.P.; Borodovsky, M.; Klenk, H.P.; Fraser, C.M.; Smith, H.O.; Woese, C
A;Title: Complete genome sequence of the methanogenic archaeon, Methanococcus jannaschii
A;Reference number: A64300; MUID:96337999; PMID:8688087
A;Accession: C64445
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-318 <BUL>
A;Cross-references: UNIPROT:Q58564; GB:U67558; GB:L77117; NID:g1591786; PIDN:AAB99166.1;
C;Genetics:
A;Map position: REV1106050-1105094
A;Start codon: GTG
C;Superfamily: Methanococcus jannaschii conserved hypothetical protein MJ1164

  Query Match             32.4%;  Score 44;  DB 2;  Length 318;
  Best Local Similarity   38.9%;  Pred. No. 90;
  Matches    7;  Conservative    5;  Mismatches    6;  Indels    0;  Gaps    0;

Qy          1 VYVFSTEMANKAAEAVLK 18
              :| : |:||| : :  ||
Db        156 IYKYETQMANPSVDVALK 173
```

```
RESULT 30
T19105
phosphate carrier protein - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
C;Accession: T19105; S44093; T20489
R;Harris, B.
submitted to the EMBL Data Library, May 1996
A;Reference number: Z19073
A;Accession: T19105
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-340 <WIL>
A;Cross-references: UNIPROT:P40614; EMBL:Z73103; PIDN:CAA97430.1; GSPDB:GN00022; CESP:F0
A;Experimental source: clone C08F8
R;Runswick, M.J.; Philippides, A.; Lauria, G.; Walker, J.E.
submitted to the EMBL Data Library, November 1993
A;Description: Extension of the mitochondrial transport superfamily: sequences of five m
A;Reference number: S44090
A;Accession: S44093
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-340 <RUN>
A;Cross-references: EMBL:X76113; NID:g472905; PID:g472906
R;Harris, B.
submitted to the EMBL Data Library, January 1996
A;Reference number: Z19281
A;Accession: T20489
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-340 <WI2>
A;Cross-references: EMBL:Z68341; PIDN:CAA92769.1; GSPDB:GN00022; CESP:F01G4.6
A;Experimental source: clone F01G4
C;Genetics:
A;Gene: CESP:F01G4.6
A;Map position: 4
A;Introns: 32/1; 72/3; 315/3
C;Superfamily: ADP,ATP carrier protein; ADP,ATP carrier protein repeat homology
C;Keywords: mitochondrion; transmembrane protein
F;137-223/Domain: ADP,ATP carrier protein repeat homology <ACR>

  Query Match             32.4%;  Score 44;  DB 2;  Length 340;
  Best Local Similarity   33.3%;  Pred. No. 97;
  Matches    8;  Conservative    7;  Mismatches    9;  Indels    0;  Gaps    0;

Qy          4 FSTEMANKAAEAVLKGQVETIVSF 27
              | | :| : | |::||   |:: :
Db         85 FRTTIAEEGARALVKGWAPTLLGY 108


RESULT 31
A85056
probable transposon protein [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 09-Jul-2004
C;Accession: A85056
R;anonymous, The European Union Arabidopsis Genome Sequencing Consortium, The Cold Sprin
Nature 402, 769-777, 1999
A;Title: Sequence and analysis of chromosome 4 of the plant Arabidopsis thaliana.
A;Reference number: A85001; MUID:20083488; PMID:10617198
A;Accession: A85056
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-350 <STO>
A;Cross-references: UNIPROT:Q9XEC1; GB:NC_001268; NID:g7267200; PIDN:CAB77911.1; GSPDB:G
C;Genetics:
A;Gene: AT4g04430
A;Map position: 4
C;Superfamily: Arabidopsis hypothetical protein F7N22.18

  Query Match             32.4%;  Score 44;  DB 2;  Length 350;
  Best Local Similarity   44.0%;  Pred. No. 1e+02;
  Matches   11;  Conservative    6;  Mismatches    6;  Indels    2;  Gaps    1;

Qy          3 VFSTEMANKAAEAVLKGQVETIVSF 27
              :|: ::|  |||| :: | | | ||
Db        260 IFTEKLA--AAEACIQSQAERINSF 282


RESULT 32
AH3455
acriflavin resistance protein A precursor [imported] - Brucella melitensis (strain 16M)
C;Species: Brucella melitensis
C;Date: 01-Feb-2002 #sequence_revision 01-Feb-2002 #text_change 09-Jul-2004
C;Accession: AH3455
R;DelVecchio, V.G.; Kapatral, V.; Redkar, R.J.; Patra, G.; Mujer, C.; Los, T.; Ivanova, N
.; Mazur, M.; Goltsman, E.; Selkov, E.; Elzer, P.H.; Hagius, S.; O'Callaghan, D.; Letessc
Proc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002
A;Title: The genome sequence of the facultative intracellular pathogen Brucella melitens
A;Reference number: AD3252; PMID:11756688
A;Accession: AH3455
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-395 <KUR>
A;Cross-references: UNIPROT:Q8YF92; GB:AE008917; PIDN:AAL52811.1; PID:g17983649; GSPDB:GN
A;Experimental source: strain 16M
C;Genetics:
A;Gene: BMEI1630
A;Map position: I

  Query Match             32.4%;  Score 44;  DB 2;  Length 395;
  Best Local Similarity   50.0%;  Pred. No. 1.1e+02;
  Matches   10;  Conservative    2;  Mismatches    8;  Indels    0;  Gaps    0;

Qy          4 FSTEMANKAAEAVLKGQVET 23
              |:|   |||| |  | |: |
Db        144 FATRDLNKAAVAAAKAQLRT 163


RESULT 33
A72498
probable DNA/pantothenate metabolism flavoprotein APE1959 - Aeropyrum pernix (strain K1)
C;Species: Aeropyrum pernix
C;Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 09-Jul-2004
C;Accession: A72498
R;Kawarabayasi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Haikawa, Y.; Jin-no, K.; Takaha
awa, H.; Takamiya, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Ku
DNA Res. 6, 83-101, 1999
A;Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, Aeropyru
A;Reference number: A72450; MUID:99310339; PMID:10382966
A;Accession: A72498
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-437 <KAW>
A;Cross-references: UNIPROT:Q9YAI0; DDBJ:AP000063; NID:g5105654; PIDN:BAA80969.1; PID:g51
A;Experimental source: strain K1
C;Genetics:
A;Gene: APE1959
C;Superfamily: pantothenate metabolism flavoprotein dfp

  Query Match             32.4%;  Score 44;  DB 2;  Length 437;
  Best Local Similarity   52.4%;  Pred. No. 1.3e+02;
  Matches   11;  Conservative    2;  Mismatches    2;  Indels    6;  Gaps    1;

Qy          8 MANKAAEAVLKGQVETIVSFH 28
              | :|: ||||||      |||
Db        405 MLDKSGEAVLKG------SFH 419


RESULT 34
AG1051
probable transport protein SgaT sgaT [imported] - Salmonella enterica subsp. enterica ser
C;Species: Salmonella enterica subsp. enterica serovar Typhi
A;Note: this species has also been called Salmonella typhi
C;Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 18-Nov-2002
```

C;Accession: AG1051
R;Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher,
th, T.; Connerton, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar,
, S.; Moule, S.; O'Gaora, P.
Nature 413, 848-852, 2001
A;Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.;
A;Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serov
A;Reference number: AB0502; MUID:21534947; PMID:11677608
A;Accession: AG1051
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-476 <PAR>
A;Cross-references: GB:AL513382; PIDN:CAD06860.1; PID:g16505508; GSPDB:GN00176
C;Genetics:
A;Gene: sgaT

Query Match 32.4%; Score 44; DB 2; Length 476;
Best Local Similarity 30.0%; Pred. No. 1.4e+02;
Matches 6; Conservative 7; Mismatches 7; Indels 0; Gaps 0;

Qy 8 MANKAAEAVLKGQVETIVSF 27
: |: ::|| ::||: |
Db 44 LLRKSVSVIIKGTIKTIIGF 63

RESULT 35
D65230
hypothetical 52.9 kD protein in aidB-rpsF intergenic region - Escherichia coli (strain K
N;Alternate names: hypothetical protein o488
C;Species: Escherichia coli
C;Date: 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change 09-Jul-2004
C;Accession: D65230; S56418
R;Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Co
.A.; Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A;Title: The complete genome sequence of Escherichia coli K-12.
A;Reference number: A64720; MUID:97426617; PMID:9278503
A;Accession: D65230
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-484 <BLAT>
A;Cross-references: UNIPROT:P39301; GB:AE000491; GB:U00096; NID:g2367357; PIDN:AAC77150.
A;Experimental source: strain K-12, substrain MG1655
R;Burland, V.; Plunkett III, G.; Sofia, H.J.; Daniels, D.L.; Blattner, F.R.
Nucleic Acids Res. 23, 2105-2119, 1995
A;Title: Analysis of the Escherichia coli genome VI: DNA sequence of the region from 92.
A;Reference number: S56314; MUID:95334362; PMID:7610040
A;Accession: S56418
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-469,'AQKKMQKNNWQNSLLNKEF' <BUR>
A;Cross-references: EMBL:U14003; NID:g1263172; PIDN:AAA97089.1; PID:g537034
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, August 1994
C;Genetics:
A;Gene: yjfS
A;Start codon: GTG

Query Match 32.4%; Score 44; DB 2; Length 484;
Best Local Similarity 30.0%; Pred. No. 1.4e+02;
Matches 6; Conservative 7; Mismatches 7; Indels 0; Gaps 0;

Qy 8 MANKAAEAVLKGQVETIVSF 27
: |: ::|| ::||: |
Db 52 LLRKSVSVIIKGTIKTIIGF 71

RESULT 36
A98275
hypothetical protein ECs5169 [imported] - Escherichia coli (strain O157:H7, substrain RI
C;Species: Escherichia coli
C;Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 09-Jul-2004
C;Accession: A98275
R;Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.
gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001
A;Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and geno
A;Reference number: A99629; MUID:21156231; PMID:11258796
A;Accession: A98275
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-484 <HAY>
A;Cross-references: UNIPROT:Q8XDJ5; GB:BA000007; PIDN:BAB38592.1; PID:g13364646; GSPDB:G
A;Experimental source: strain O157:H7, substrain RIMD 0509952
C;Genetics:
A;Gene: ECs5169

Query Match 32.4%; Score 44; DB 2; Length 484;
Best Local Similarity 30.0%; Pred. No. 1.4e+02;
Matches 6; Conservative 7; Mismatches 7; Indels 0; Gaps 0;

Qy 8 MANKAAEAVLKGQVETIVSF 27
: |: ::|| ::||: |
Db 52 LLRKSVSVIIKGTIKTIIGF 71

RESULT 37
A86116
hypothetical protein sgaT [imported] - Escherichia coli (strain O157:H7, substrain EDL93
C;Species: Escherichia coli
C;Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 09-Jul-2004
C;Accession: A86116
R;Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
iller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,
Nature 409, 529-533, 2001
A;Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A;Reference number: A85480; MUID:21074935; PMID:11206551
A;Accession: A86116
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-484 <STO>
A;Cross-references: UNIPROT:Q8XDJ5; GB:AE005174; NID:g12519184; PIDN:AAG59389.1; GSPDB:G
A;Experimental source: strain O157:H7, substrain EDL933
C;Genetics:
A;Gene: sgaT

Query Match 32.4%; Score 44; DB 2; Length 484;
Best Local Similarity 30.0%; Pred. No. 1.4e+02;
Matches 6; Conservative 7; Mismatches 7; Indels 0; Gaps 0;

Qy 8 MANKAAEAVLKGQVETIVSF 27
: |: ::|| ::||: |
Db 52 LLRKSVSVIIKGTIKTIIGF 71

RESULT 38
T05092
probable 1,2-diacylglycerol 3-beta-galactosyltransferase (EC 2.4.1.46) - Arabidopsis tha
N;Alternate names: protein F28M20.30
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 09-Jul-2004
C;Accession: T05092
R;Bevan, M.; Rieger, M.; Mueller-Auer, S.; Zipp, M.; Schaefer, M.; Hoheisel, J.; Mewes,
submitted to the Protein Sequence Database, November 1998
A;Reference number: Z15398
A;Accession: T05092
A;Molecule type: DNA
A;Residues: 1-533 <BEV>
A;Cross-references: UNIPROT:Q9MU68; UNIPROT:O81770; EMBL:AL031004
A;Experimental source: cultivar Columbia; BAC clone F28M20
C;Genetics:
A;Map position: 4
A;Introns: 175/3; 233/2; 287/2; 334/3; 409/3; 430/3; 455/3
A;Note: F28M20.30
C;Keywords: glycosyltransferase; hexosyltransferase

Query Match 32.4%; Score 44; DB 2; Length 533;

```
  Best Local Similarity   37.5%;  Pred. No. 1.6e+02;
  Matches    9;  Conservative    7;  Mismatches    8;  Indels    0;  Gaps    0;

Qy          5 STEMANKAAEAVLKGQVETIVSFH 28
              || :| : |: ::| | : |:| |
Db        228 STFIAREIAQGLMKYQPDIIISVH 251


RESULT 39
JC5756
vibriolysin (EC 3.4.24.-) precursor - Vibrio vulnificus
N;Alternate names: zinc metalloproteinase
C;Species: Vibrio vulnificus
C;Date: 24-Jan-1998 #sequence_revision 13-Mar-1998 #text_change 09-Jul-2004
C;Accession: JC5756; PC4496
R;Cheng, J.C.; Shao, C.P.; Hor, L.I.
Gene 183, 255-257, 1996
A;Title: Cloning and nucleotide sequencing of the protease gene of Vibrio vulnificus.
A;Reference number: JC5756; MUID:97149307; PMID:8996115
A;Accession: JC5756
A;Molecule type: DNA
A;Residues: 1-609 <CHE>
A;Cross-references: UNIPROT:P96176; GB:U48780; NID:g1794193; PIDN:AAC44789.1; PID:g17941
A;Experimental source: strain YJ016
A;Accession: PC4496
A;Molecule type: protein
A;Residues: 197-216 <CH2>
C;Genetics:
A;Gene: uup
C;Superfamily: zinc metalloendopeptidase, neutral protease type (elastase)
C;Keywords: hydrolase; metalloproteinase; zinc
F;1-25/Domain: signal sequence #status predicted <SIG>
F;26-197/Domain: propeptide #status predicted <PRO>
F;197-609/Product: vibriolysin #status predicted <MAT>
F;231-257,473-502/Disulfide bonds: #status predicted
F;343,347,367/Binding site: zinc (His, His, Glu) #status predicted
F;344,426/Active site: Glu, His #status predicted

  Query Match            32.4%;  Score 44;  DB 2;  Length 609;
  Best Local Similarity   45.0%;  Pred. No. 1.8e+02;
  Matches    9;  Conservative    6;  Mismatches    5;  Indels    0;  Gaps    0;

Qy          6 TEMANKAAEAVLKGQVETIV 25
              :::| :|||  :|| |: ||
Db        370 SDIAGEAAEFYMKGSVDWIV 389


RESULT 40
G72709
probable DNA ligase APE1094 - Aeropyrum pernix (strain K1)
C;Species: Aeropyrum pernix
C;Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Jun-2000
C;Accession: G72709
R;Kawarabayasi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Haikawa, Y.; Jin-no, K.; Takah
awa, H.; Takamiya, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; K
DNA Res. 6, 83-101, 1999
A;Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, Aeropyr
A;Reference number: A72450; MUID:99310339; PMID:10382966
A;Accession: G72709
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-619 <KAW>
A;Cross-references: DDBJ:AP000060; NID:g5104188; PIDN:BAA80079.1; PID:g5104764
A;Experimental source: strain K1
C;Genetics:
A;Gene: APE1094
C;Superfamily: DNA ligase

  Query Match            32.4%;  Score 44;  DB 2;  Length 619;
  Best Local Similarity   22.5%;  Pred. No. 1.8e+02;
  Matches    9;  Conservative    9;  Mismatches    6;  Indels   16;  Gaps    1;

Qy          1 VYVFSTEMAN----------------KAAEAVLKGQVETI 24
                ||::|  : |                || ||:::|::  :
Db        296 VYIYSRRLENITRMFPDVVEMARKGLKAGEAIVEGEIVAV 335


Search completed: June  8, 2005, 03:23:53
Job time : 25.4375 secs
```

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: June 8, 2005, 03:00:28 ; Search time 127.75 Seconds
(without alignments)
112.237 Million cell updates/sec

Title: US-09-915-543-15_COPY_177_204
Perfect score: 136
Sequence: 1 VYVFSTEMANKAAEAVLKGQVETIVSFH 28

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1612378 seqs, 512079187 residues

Total number of hits satisfying chosen parameters: 1612378

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : UniProt_03:*
1: uniprot_sprot:*
2: uniprot_trembl:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	136	100.0	1425	2	Q67FX9	Q67fx9 mus musculu
2	136	100.0	1426	1	BCL9_HUMAN	O00512 homo sapien
3	122	89.7	796	2	Q6NRE2	Q6nre2 xenopus lae
4	102	75.0	1474	2	Q67FY0	Q67fy0 brachydanio
5	98	72.1	1469	1	BCL9_DROME	Q961d9 drosophila
6	94	69.1	1457	2	Q641L9	Q641l9 mus musculu
7	94	69.1	1494	2	Q67FY2	Q67fy2 mus musculu
8	94	69.1	1494	2	Q6I7B5	Q6i7b5 mus musculu
9	91	66.9	1494	2	Q67FY1	Q67fy1 homo sapien
10	91	66.9	1499	2	Q86UU0	Q86uu0 homo sapien
11	84	61.8	1530	2	Q67FY3	Q67fy3 brachydanio
12	60	44.1	205	1	HEM2_CLOJO	Q59295 clostridium
13	55	40.4	1050	2	Q9U1Q4	Q9u1q4 caenorhabdi
14	55	40.4	1062	2	Q767M3	Q767m3 sus scrofa
15	54	39.7	193	2	Q6HQ86	Q6hq86 bacillus an
16	54	39.7	523	2	Q630I9	Q630i9 bacillus ce
17	54	39.7	523	2	Q72X37	Q72x37 bacillus ce
18	54	39.7	523	2	Q6HAM4	Q6ham4 bacillus th
19	53	39.0	516	2	Q71SH3	Q71sh3 oryza sativ
20	53	39.0	516	2	Q7XN11	Q7xn11 oryza sativ
21	52	38.2	1059	2	Q6ZQL7	Q6zql7 mus musculu
22	52	38.2	1065	2	Q6MG21	Q6mg21 rattus norv
23	52	38.2	1086	2	Q69Z78	Q69z78 mus musculu
24	51	37.5	243	2	Q97JA0	Q97ja0 clostridium
25	51	37.5	520	2	Q84P52	Q84p52 lycopersico
26	51	37.5	1049	2	Q960E6	Q960e6 drosophila
27	51	37.5	1049	2	Q9V6L1	Q9v6l1 drosophila
28	51	37.5	1217	1	SYV_FUGRU	P49696 fugu rubrip
29	50	36.8	338	2	Q89JQ9	Q89jq9 bradyrhizob
30	50	36.8	459	2	O82521	O82521 capsicum ch
31	50	36.8	523	2	Q814L6	Q814l6 bacillus ce
32	50	36.8	642	2	Q9H6R2	Q9h6r2 homo sapien
33	50	36.8	657	2	Q6DKJ5	Q6dkj5 homo sapien
34	50	36.8	733	2	Q96GN2	Q96gn2 homo sapien
35	50	36.8	1098	2	Q96Q02	Q96q02 homo sapien
36	49.5	36.4	350	2	Q8GGI9	Q8ggi9 lactobacill
37	49	36.0	141	2	Q9LFW7	Q9lfw7 arabidopsis
38	49	36.0	143	2	Q976C9	Q976c9 sulfolobus
39	49	36.0	187	2	Q8G3S3	Q8g3s3 bifidobacte
40	49	36.0	210	2	Q9FZL1	Q9fzl1 arabidopsis
41	49	36.0	233	2	Q96YA9	Q96ya9 sulfolobus
42	49	36.0	237	2	Q96X55	Q96x55 sulfolobus
43	49	36.0	268	2	Q8DFX2	Q8dfx2 vibrio vuln
44	49	36.0	446	2	Q6C2I5	Q6c2i5 yarrowia li
45	49	36.0	458	2	O83532	O83532 treponema p

ALIGNMENTS

RESULT 1
Q67FX9
ID Q67FX9 PRELIMINARY; PRT; 1425 AA.
AC Q67FX9;
DT 25-OCT-2004 (TrEMBLrel. 28, Created)
DT 25-OCT-2004 (TrEMBLrel. 28, Last sequence update)
DT 25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DE BCL9.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6;
RA Brembeck F.H., Schwarz-Romond T., Bakkers J., Wilhelm S.,
RA Hammerschmidt M., Birchmeier W.;
RT "Essential role of BCL9-2 in the switch between [beta]-catenin's
RT adhesive and transcriptional functions.";
RL Genes Dev. 18:0-0(2004).
DR EMBL; AY296061; AAQ62699.1; -.
SQ SEQUENCE 1425 AA; 148970 MW; 77347CF56FC4A815 CRC64;

Query Match 100.0%; Score 136; DB 2; Length 1425;
Best Local Similarity 100.0%; Pred. No. 4.2e-11;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 VYVFSTEMANKAAEAVLKGQVETIVSFH 28
              ||||||||||||||||||||||||||||
Db        177 VYVFSTEMANKAAEAVLKGQVETIVSFH 204
```

RESULT 2
BCL9_HUMAN
ID BCL9_HUMAN STANDARD; PRT; 1426 AA.
AC O00512;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 05-JUL-2004 (Rel. 44, Last annotation update)
DE B-cell lymphoma 9 protein (Bcl-9) (Legless homolog).
GN Name=BCL9;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Fetal brain;
RX MEDLINE=98158621; PubMed=9490669;
RA Willis T.G., Zalcberg I.R., Coignet L.J.A., Wlodarska I., Stul M.,
RA Jadayel D.M., Bastard C., Treleaven J.G., Catovsky D., Silva M.L.M.,
RA Dyer M.J.S.;
RT "Molecular cloning of translocation t(1;14)(q21;q32) defines a novel
RT gene (BCL9) at chromosome 1q21.";

```
RL   Blood 91:1873-1881(1998).
RN   [2]
RP   FUNCTION.
RX   MEDLINE=21952490; PubMed=11955446; DOI=10.1016/S0092-8674(02)00679-7;
RA   Kramps T., Peter O., Brunner E., Nellen D., Froesch B., Chatterjee S.,
RA   Murone M., Zuellig S., Basler K.;
RT   "Wnt/wingless signaling requires BCL9/legless-mediated recruitment of
RT   pygopus to the nuclear beta-catenin-TCF complex.";
RL   Cell 109:47-60(2002).
CC   -!- FUNCTION: Involved in signal transduction through the Wnt pathway.
CC   -!- SUBUNIT: Binds to beta-catenin (CTNNB1), PYGO1 and PYGO2.
CC   -!- SUBCELLULAR LOCATION: Nuclear (Probable).
CC   -!- TISSUE SPECIFICITY: Detected at low levels in thymus, prostate,
CC       testis, ovary and small intestine, and at lower levels in spleen,
CC       colon and blood.
CC   -!- DISEASE: Involved in a t(1;14)(q21;q32) chromosomal translocation
CC       found in a patient with precusor B-cell acute lymphoblastic
CC       leukemia (ALL). This translocation leaves the coding region
CC       intact, but may have pathogenic effects due to alterations in the
CC       expression level of BCL9. Several cases of translocations within
CC       the 3' untranslated region of BCL9 have been found in B-cell
CC       malignancies.
CC   -!- CAUTION: It is uncertain whether Met-1 or Met-27 is the initiator.
CC   -!- CAUTION: Ref.1 sequence differs from that shown due to a
CC       frameshift in position 1391.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Y13620; CAA73942.1; ALT_FRAME.
DR   Genew; HGNC:1008; BCL9.
DR   MIM; 602597; -.
KW   Chromosomal translocation; Nuclear protein; Proto-oncogene;
KW   Wnt signaling pathway.
FT   DOMAIN       231   1378       Pro-rich.
FT   DOMAIN       347    377       CTNNB1-binding.
FT   DOMAIN       331    335       Poly-Pro.
FT   DOMAIN       514    517       Poly-Pro.
FT   DOMAIN       900    903       Poly-Ala.
FT   DOMAIN       970    973       Poly-Pro.
SQ   SEQUENCE   1426 AA;  149314 MW;  A240A487716B7F1B CRC64;

  Query Match             100.0%;  Score 136;  DB 1;  Length 1426;
  Best Local Similarity   100.0%;  Pred. No. 4.2e-11;
  Matches   28;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 VYVFSTEMANKAAEAVLKGQVETIVSFH 28
              ||||||||||||||||||||||||||||
Db        177 VYVFSTEMANKAAEAVLKGQVETIVSFH 204


RESULT 3
Q6NRE2
ID   Q6NRE2      PRELIMINARY;      PRT;   796 AA.
AC   Q6NRE2;
DT   05-JUL-2004 (TrEMBLrel. 27, Created)
DT   05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT   05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE   MGC83888 protein.
GN   Name=MGC83888;
OS   Xenopus laevis (African clawed frog).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
OC   Xenopodinae; Xenopus.
OX   NCBI_TaxID=8355;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Oocytes;
RX   MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA   Krzywinski M.I., Skalska U., Smailus D.E., Schnerch A., Schein J.E.,
RA   Jones S.J., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length human
RT   and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Oocytes;
RX   MEDLINE=22341132; PubMed=12454917; DOI=10.1002/dvdy.10174;
RA   Klein S.L., Strausberg R.L., Wagner L., Pontius J., Clifton S.W.,
RA   Richardson P.;
RT   "Genetic and genomic tools for Xenopus research: The NIH Xenopus
RT   initiative.";
RL   Dev. Dyn. 225:384-391(2002).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Oocytes;
RA   Klein S., Strausberg R.;
RL   Submitted (MAY-2004) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC070813; AAH70813.1; -.
SQ   SEQUENCE   796 AA;  86048 MW;  9A282C1DCA316678 CRC64;

  Query Match             89.7%;  Score 122;  DB 2;  Length 796;
  Best Local Similarity   89.3%;  Pred. No. 2.9e-09;
  Matches   25;  Conservative    1;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 VYVFSTEMANKAAEAVLKGQVETIVSFH 28
              |||||||||||||||||||| |:|| ||
Db        173 VYVFSTEMANKAAEAVLKGQAESIVLFH 200


RESULT 4
Q67FY0
ID   Q67FY0      PRELIMINARY;      PRT;  1474 AA.
AC   Q67FY0;
DT   25-OCT-2004 (TrEMBLrel. 28, Created)
DT   25-OCT-2004 (TrEMBLrel. 28, Last sequence update)
DT   25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DE   Bcl9.
OS   Brachydanio rerio (Zebrafish) (Danio rerio).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC   Cyprinidae; Danio.
OX   NCBI_TaxID=7955;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Brembeck F.H., Schwarz-Romond T., Bakkers J., Wilhelm S.,
RA   Hammerschmidt M., Birchmeier W.;
RT   "Essential role of BCL9-2 in the switch between [beta]-catenin's
RT   adhesive and transcriptional functions.";
RL   Genes Dev. 18:0-0(2004).
DR   EMBL; AY296060; AAQ62698.1; -.
SQ   SEQUENCE   1474 AA;  154339 MW;  4B2C3E8092BE3532 CRC64;

  Query Match             75.0%;  Score 102;  DB 2;  Length 1474;
  Best Local Similarity   64.3%;  Pred. No. 5.3e-06;
  Matches   18;  Conservative    7;  Mismatches    3;  Indels    0;  Gaps    0;
```

```
Qy          1 VYVFSTEMANKAAEAVLKGQVETIVSFH 28
              ||||:||||||||:||: | |: |:::|
Db        208 VYVFTTEMANKAADAVITGHVDNIITYH 235


RESULT 5
BCL9_DROME
ID   BCL9_DROME     STANDARD;      PRT;  1469 AA.
AC   Q961D9; Q9V4D2;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   25-JAN-2005 (Rel. 46, Last annotation update)
DE   Bcl-9 homolog (Legless protein).
GN   Name=lgs; Synonyms=BCL9; ORFNames=CG2041;
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC   Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Berkeley;
RX   MEDLINE=20196006; PubMed=10731132; DOI=10.1126/science.287.5461.2185;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D.A., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J.H., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
RN   [2]
RP   GENOME REANNOTATION.
RX   MEDLINE=22426069; PubMed=12537572;
RA   Misra S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,
RA   Hradecky P., Huang Y., Kaminker J.S., Millburn G.H., Prochnik S.E.,
RA   Smith C.D., Tupy J.L., Whitfield E.J., Bayraktaroglu L., Berman B.P.,
RA   Bettencourt B.R., Celniker S.E., de Grey A.D.N.J., Drysdale R.A.,
RA   Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.Q.,
RA   Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,
RA   Lewis S.E.;
RT   "Annotation of the Drosophila melanogaster euchromatic genome: a
RT   systematic review.";
RL   Genome Biol. 3:RESEARCH0083.1-RESEARCH0083.22(2002).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Berkeley; TISSUE=Embryo;
RX   MEDLINE=22426066; PubMed=12537569;
RA   Stapleton M., Carlson J.W., Brokstein P., Yu C., Champe M.,
RA   George R.A., Guarin H., Kronmiller B., Pacleb J.M., Park S., Wan K.H.,
RA   Rubin G.M., Celniker S.E.;
RT   "A Drosophila full-length cDNA resource.";
RL   Genome Biol. 3:RESEARCH0080.1-RESEARCH0080.8(2002).
RN   [4]
RP   SEQUENCE OF 6-1469 FROM N.A., AND MUTAGENESIS OF GLY-514; LEU-534 AND
RP   ILE-537.
RX   MEDLINE=21952490; PubMed=11955446; DOI=10.1016/S0092-8674(02)00679-7;
RA   Kramps T., Peter O., Brunner E., Nellen D., Froesch B., Chatterjee S.,
RA   Murone M., Zuellig S., Basler K.;
RT   "Wnt/wingless signaling requires BCL9/legless-mediated recruitment of
RT   pygopus to the nuclear beta-catenin-TCF complex.";
RL   Cell 109:47-60(2002).
CC   -!- FUNCTION: Involved in signal transduction through the Wnt pathway.
CC   -!- SUBUNIT: Binds to ARM and PYGO.
CC   -!- SUBCELLULAR LOCATION: Nuclear.
CC   -!- DEVELOPMENTAL STAGE: Expressed both maternally and zygotically
CC       throughout development.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AE003844; AAF59345.2; -.
DR   EMBL; AY051651; AAK93075.1; -.
DR   EMBL; AF457205; AAL91368.1; -.
DR   FlyBase; FBgn0039907; lgs.
DR   GO; GO:0005634; C:nucleus; IDA.
DR   GO; GO:0030528; F:transcription regulator activity; IPI.
DR   GO; GO:0030177; P:positive regulation of Wnt receptor signali. . .; IPI.
DR   GO; GO:0007367; P:segment polarity determination; IMP.
KW   Developmental protein; Nuclear protein; Segmentation polarity protein;
KW   Wnt signaling pathway.
FT   DOMAIN      511    555       ARM-binding.
FT   DOMAIN     1134   1173       Asn-rich.
FT   DOMAIN     1340   1449       Gln-rich.
FT   DOMAIN     1162   1169       Poly-Asn.
FT   MUTAGEN     514    514       G->E: In allele lgs-21L.
FT   MUTAGEN     534    534       L->F: In allele lgs-17E; segment polarity
FT                                phenotype.
FT   MUTAGEN     537    537       I->K: In allele lgs-17P.
SQ   SEQUENCE   1469 AA;  153759 MW;  5672E01B7200ED08 CRC64;

  Query Match             72.1%;  Score 98;  DB 1;  Length 1469;
  Best Local Similarity   57.1%;  Pred. No. 2.1e-05;
  Matches   16;  Conservative    9;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          1 VYVFSTEMANKAAEAVLKGQVETIVSFH 28
              ::||||::||| ||:|| || :||:::|
Db        323 IFVFSTQLANKGAESVLSGQFQTIIAYH 350


RESULT 6
Q641L9
ID   Q641L9      PRELIMINARY;      PRT;  1457 AA.
AC   Q641L9;
DT   25-OCT-2004 (TrEMBLrel. 28, Created)
DT   25-OCT-2004 (TrEMBLrel. 28, Last sequence update)
DT   25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DE   Bcl9l protein.
GN   Name=Bcl9l;
OS   Mus musculus (Mouse).
```

```
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6; TISSUE=Brain;
RX   PubMed=12477932; DOI=10.1073/pnas.242603899;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA   Krzywinski M.I., Skalska U., Smailus D.E., Schnerch A., Schein J.E.,
RA   Jones S.J., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length human
RT   and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6; TISSUE=Brain;
RA   Director MGC Project;
RL   Submitted (SEP-2004) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC082304; AAH82304.1; -.
SQ   SEQUENCE   1457 AA;  152636 MW;  4FD2B47ADDE92A33 CRC64;

  Query Match             69.1%;  Score 94;  DB 2;  Length 1457;
  Best Local Similarity   60.7%;  Pred. No. 8.2e-05;
  Matches   17;  Conservative    8;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          1 VYVFSTEMANKAAEAVLKGQVETIVSFH 28
              ||||:| :|| ||||||:|: |:|:::|
Db        201 VYVFTTHLANTAAEAVLQGRAESILAYH 228


RESULT 7
Q67FY2
ID   Q67FY2      PRELIMINARY;      PRT;  1494 AA.
AC   Q67FY2;
DT   25-OCT-2004 (TrEMBLrel. 28, Created)
DT   25-OCT-2004 (TrEMBLrel. 28, Last sequence update)
DT   25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DE   BCL9-2.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6;
RA   Brembeck F.H., Schwarz-Romond T., Bakkers J., Wilhelm S.,
RA   Hammerschmidt M., Birchmeier W.;
RT   "Essential role of BCL9-2 in the switch between [beta]-catenin's
RT   adhesive and transcriptional functions.";
RL   Genes Dev. 18:0-0(2004).
DR   EMBL; AY296058; AAQ62696.1; -.
SQ   SEQUENCE   1494 AA;  156679 MW;  31A9904C5923581C CRC64;

  Query Match             69.1%;  Score 94;  DB 2;  Length 1494;
  Best Local Similarity   60.7%;  Pred. No. 8.4e-05;
  Matches   17;  Conservative    8;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          1 VYVFSTEMANKAAEAVLKGQVETIVSFH 28
              ||||:| :|| ||||||:|: |:|:::|
Db        238 VYVFTTHLANTAAEAVLQGRAESILAYH 265


RESULT 8
Q6I7B5
ID   Q6I7B5      PRELIMINARY;      PRT;  1494 AA.
AC   Q6I7B5;
DT   05-JUL-2004 (TrEMBLrel. 27, Created)
DT   05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT   05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE   BCL9-related beta-catenin-binding protein.
GN   Name=B9L;
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Adachi S., Jigami T., Yasui T., Nakano T., Ohwada S., Omori Y.,
RA   Sugano S., Ohkawara B., Shibuya H., Nakamura T., Akiyama T.;
RL   Submitted (DEC-2003) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB128033; BAD24964.1; -.
SQ   SEQUENCE   1494 AA;  156570 MW;  71AE96FD33743A6C CRC64;

  Query Match             69.1%;  Score 94;  DB 2;  Length 1494;
  Best Local Similarity   60.7%;  Pred. No. 8.4e-05;
  Matches   17;  Conservative    8;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          1 VYVFSTEMANKAAEAVLKGQVETIVSFH 28
              ||||:| :|| ||||||:|: |:|:::|
Db        238 VYVFTTHLANTAAEAVLQGRAESILAYH 265


RESULT 9
Q67FY1
ID   Q67FY1      PRELIMINARY;      PRT;  1494 AA.
AC   Q67FY1;
DT   25-OCT-2004 (TrEMBLrel. 28, Created)
DT   25-OCT-2004 (TrEMBLrel. 28, Last sequence update)
DT   25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DE   BCL9-2.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Brembeck F.H., Schwarz-Romond T., Bakkers J., Wilhelm S.,
RA   Hammerschmidt M., Birchmeier W.;
RT   "Essential role of BCL9-2 in the switch between [beta]-catenin's
RT   adhesive and transcriptional functions.";
RL   Genes Dev. 18:0-0(2004).
DR   EMBL; AY296059; AAQ62697.1; -.
SQ   SEQUENCE   1494 AA;  156528 MW;  2D591F45FF3AEE36 CRC64;

  Query Match             66.9%;  Score 91;  DB 2;  Length 1494;
  Best Local Similarity   57.1%;  Pred. No. 0.00024;
  Matches   16;  Conservative    9;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          1 VYVFSTEMANKAAEAVLKGQVETIVSFH 28
              ||||:| :|| ||||||:|: ::|:::|
Db        235 VYVFTTHLANTAAEAVLQGRADSILAYH 262


RESULT 10
Q86UU0
ID   Q86UU0      PRELIMINARY;      PRT;  1499 AA.
AC   Q86UU0;
DT   01-JUN-2003 (TrEMBLrel. 24, Created)
DT   01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   DLNB11 protein.
GN   Name=DLNB11;
```

```
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Kubo T., Arai Y., Ohira M., Gamou T., Maeno G., Sakiyama T.,
RA   Toyoda A., Hattori M., Sakaki Y., Nakagawara A., Ohki M.;
RL   Submitted (OCT-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB094091; BAC76045.1; -.
SQ   SEQUENCE   1499 AA;  157129 MW;  8415C2EDB7AA9C0C CRC64;

  Query Match             66.9%;  Score 91;  DB 2;  Length 1499;
  Best Local Similarity   57.1%;  Pred. No. 0.00024;
  Matches   16;  Conservative    9;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          1 VYVFSTEMANKAAEAVLKGQVETIVSFH 28
              ||||:| :|| ||||||:|: ::|:::|
Db        240 VYVFTTHLANTAAEAVLQGRADSILAYH 267


RESULT 11
Q67FY3
ID   Q67FY3      PRELIMINARY;      PRT;  1530 AA.
AC   Q67FY3;
DT   25-OCT-2004 (TrEMBLrel. 28, Created)
DT   25-OCT-2004 (TrEMBLrel. 28, Last sequence update)
DT   25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DE   Bcl9-2.
OS   Brachydanio rerio (Zebrafish) (Danio rerio).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC   Cyprinidae; Danio.
OX   NCBI_TaxID=7955;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Brembeck F.H., Schwarz-Romond T., Bakkers J., Wilhelm S.,
RA   Hammerschmidt M., Birchmeier W.;
RT   "Essential role of BCL9-2 in the switch between [beta]-catenin's
RT   adhesive and transcriptional functions.";
RL   Genes Dev. 18:0-0(2004).
DR   EMBL; AY296057; AAQ62695.1; -.
SQ   SEQUENCE   1530 AA;  159872 MW;  C29FEC9433ED28C0 CRC64;

  Query Match             61.8%;  Score 84;  DB 2;  Length 1530;
  Best Local Similarity   53.6%;  Pred. No. 0.0027;
  Matches   15;  Conservative    7;  Mismatches    6;  Indels    0;  Gaps    0;

Qy          1 VYVFSTEMANKAAEAVLKGQVETIVSFH 28
              ||||:| :|| |||||: |  ::|: :|
Db        195 VYVFTTSLANSAAEAVMHGHTDSILLYH 222


RESULT 12
HEM2_CLOJO
ID   HEM2_CLOJO      STANDARD;      PRT;   205 AA.
AC   Q59295;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   05-JUL-2004 (Rel. 44, Last annotation update)
DE   Delta-aminolevulinic acid dehydratase (EC 4.2.1.24) (Porphobilinogen
DE   synthase) (ALAD) (ALADH) (Fragment).
GN   Name=hemB;
OS   Clostridium josui.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1499;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=FERM P-9684;
RX   MEDLINE=95394829; PubMed=7665501;
RA   Fujino E., Fujino T., Karita S., Sakka K., Ohmiya K.;
RT   "Cloning and sequencing of some genes responsible for porphyrin
RT   biosynthesis from the anaerobic bacterium Clostridium josui.";
RL   J. Bacteriol. 177:5169-5175(1995).
CC   -!- CATALYTIC ACTIVITY: 2 5-aminolevulinate = porphobilinogen + 2
CC       H(2)O.
CC   -!- COFACTOR: Zinc (By similarity).
CC   -!- PATHWAY: Siroheme biosynthesis.
CC   -!- SUBUNIT: Homooctamer (By similarity).
CC   -!- SIMILARITY: Belongs to the ALADH family.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; D28503; BAA05863.1; -.
DR   PIR; I40812; I40812.
DR   HSSP; P15002; 1L6S.
DR   InterPro; IPR001731; AlaD_dehydratase.
DR   Pfam; PF00490; ALAD; 1.
DR   PRINTS; PR00144; DALDHYDRTASE.
DR   ProDom; PD002304; AlaD_dehydratase; 1.
DR   PROSITE; PS00169; D_ALA_DEHYDRATASE; PARTIAL.
KW   Lyase; Porphyrin biosynthesis; Zinc.
FT   DOMAIN      114    132       Zinc-binding (By similarity).
FT   NON_TER     205    205
SQ   SEQUENCE   205 AA;  23172 MW;  886F9DAEFDB1144E CRC64;

  Query Match             44.1%;  Score 60;  DB 1;  Length 205;
  Best Local Similarity   46.2%;  Pred. No. 1.3;
  Matches   12;  Conservative    5;  Mismatches    9;  Indels    0;  Gaps    0;

Qy          2 YVFSTEMANKAAEAVLKGQVETIVSF 27
              | || :|  || || ||  |:::: |
Db         51 YHFSPDMVGKAIEAALKADVKSVLLF 76


RESULT 13
Q9U1Q4
ID   Q9U1Q4      PRELIMINARY;      PRT;  1050 AA.
AC   Q9U1Q4;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE   Hypothetical protein Y87G2A.5.
GN   Name=vrs-2; ORFNames=Y87G2A.5;
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Bristol N2;
RX   MEDLINE=99069613; PubMed=9851916;
RA   none;
RT   "Genome sequence of the nematode C.elegans: A platform for
RT   investigating biology.";
RL   Science 282:2012-2018(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Bristol N2;
RA   White S.;
RL   Submitted (MAY-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AL110500; CAB60428.1; -.
DR   HSSP; P96142; 1IVS.
DR   WormBase; WBGene00006936; vrs-2.
DR   WormPep; Y87G2A.5; CE24685.
DR   GO; GO:0005524; F:ATP binding; IEA.
DR   GO; GO:0004832; F:valine-tRNA ligase activity; IEA.
DR   GO; GO:0006438; P:valyl-tRNA aminoacylation; IEA.
DR   InterPro; IPR002300; tRNA-synt_1a.
```

```
DR   InterPro; IPR001412; tRNA-synt_I.
DR   InterPro; IPR002303; tRNA-synt_val.
DR   InterPro; IPR009080; tRNAsyn_1a_bind.
DR   InterPro; IPR010978; tRNA_binding_arm.
DR   InterPro; IPR009008; ValRS_IleRS_edit.
DR   Pfam; PF00133; tRNA-synt_1; 1.
DR   PRINTS; PR00986; TRNASYNTHVAL.
DR   TIGRFAMs; TIGR00422; valS; 1.
DR   PROSITE; PS00178; AA_TRNA_LIGASE_I; 1.
KW   Hypothetical protein.
SQ   SEQUENCE   1050 AA;  118920 MW;  F33DB53587EAC057 CRC64;

  Query Match             40.4%;  Score 55;  DB 2;  Length 1050;
  Best Local Similarity   44.4%;  Pred. No. 40;
  Matches   12;  Conservative    2;  Mismatches   13;  Indels    0;  Gaps    0;

Qy          2 YVFSTEMANKAAEAVLKGQVETIVSFH 28
              ||    || ||  ||  | :: |  ||
Db        461 YVKCAHMAEKAVAAVANGDLQIIPEFH 487


RESULT 14
Q767M3
ID   Q767M3      PRELIMINARY;      PRT;  1062 AA.
AC   Q767M3;
DT   05-JUL-2004 (TrEMBLrel. 27, Created)
DT   05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT   05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE   Hypothetical protein KIAA1885.
GN   Name=KIAA1885;
OS   Sus scrofa (Pig).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX   NCBI_TaxID=9823;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=larg white, and large white;
RA   Shigenari A., Ando A., Renard C., Chardon P., Shiina T., Kulski J.K.,
RA   Yasue H., Inoko H.;
RT   "Nucleotide sequencing analysis of the swine 433-kb genomic segment
RT   located between the non-classical and classical SLA class I gene
RT   clusters.";
RL   Immunogenetics 55:695-705(2004).
DR   EMBL; AB113355; BAD08425.1; -.
DR   EMBL; AB113354; BAD08423.1; -.
DR   GO; GO:0005524; F:ATP binding; IEA.
DR   GO; GO:0004832; F:valine-tRNA ligase activity; IEA.
DR   GO; GO:0006438; P:valyl-tRNA aminoacylation; IEA.
DR   InterPro; IPR002110; ANK.
DR   InterPro; IPR002300; tRNA-synt_1a.
DR   InterPro; IPR001412; tRNA-synt_I.
DR   InterPro; IPR002303; tRNA-synt_val.
DR   InterPro; IPR009080; tRNAsyn_1a_bind.
DR   InterPro; IPR009008; ValRS_IleRS_edit.
DR   Pfam; PF00133; tRNA-synt_1; 1.
DR   PRINTS; PR01415; ANKYRIN.
DR   PRINTS; PR00986; TRNASYNTHVAL.
DR   TIGRFAMs; TIGR00422; valS; 1.
DR   PROSITE; PS00178; AA_TRNA_LIGASE_I; 1.
SQ   SEQUENCE   1062 AA;  118287 MW;  619F230CC078EEC7 CRC64;

  Query Match             40.4%;  Score 55;  DB 2;  Length 1062;
  Best Local Similarity   44.4%;  Pred. No. 40;
  Matches   12;  Conservative    5;  Mismatches   10;  Indels    0;  Gaps    0;

Qy          2 YVFSTEMANKAAEAVLKGQVETIVSFH 28
              :|   ||  :||:||: | :|   |||
Db        481 FVRCQEMGEQAAKAVVSGALELSPSFH 507


RESULT 15
Q6HQ86
ID   Q6HQ86      PRELIMINARY;      PRT;  193 AA.
AC   Q6HQ86;
DT   05-JUL-2004 (TrEMBLrel. 27, Created)
DT   05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT   05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE   Two-component sensor protein, C-terminus.
GN   OrderedLocusNames=BAS5264;
OS   Bacillus anthracis.
OC   Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX   NCBI_TaxID=1392;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Sterne;
RA   Brettin T.S., Bruce D., Challacombe J.F., Gilna P., Han C., Hill K.,
RA   Hitchcock P., Jackson P., Keim P., Longmire J., Lucas S., Okinaka R.,
RA   Richardson P., Rubin E., Tice H.;
RL   Submitted (JAN-2004) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AE017225; AAT57552.1; -.
DR   GO; GO:0005524; F:ATP binding; IEA.
DR   InterPro; IPR003594; ATPbind_ATPase.
DR   Pfam; PF02518; HATPase_c; 1.
SQ   SEQUENCE   193 AA;  21865 MW;  DE1F60ACD9C3E0D0 CRC64;

  Query Match             39.7%;  Score 54;  DB 2;  Length 193;
  Best Local Similarity   54.5%;  Pred. No. 9.9;
  Matches   12;  Conservative    2;  Mismatches    8;  Indels    0;  Gaps    0;

Qy          3 VFSTEMANKAAEAVLKGQVETI 24
              :||      | |||||||| | :
Db         19 LFSLTFMTKGAEAVLKGQNEKV 40


RESULT 16
Q630I9
ID   Q630I9      PRELIMINARY;      PRT;  523 AA.
AC   Q630I9;
DT   25-OCT-2004 (TrEMBLrel. 28, Created)
DT   25-OCT-2004 (TrEMBLrel. 28, Last sequence update)
DT   25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DE   Sensor histidine kinase (EC 2.7.3.-).
GN   ORFNames=BTZK5109;
OS   Bacillus cereus ZK.
OC   Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX   NCBI_TaxID=288681;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ZK;
RA   Brettin T.S., Bruce D., Challacombe J.F., Gilna P., Han C., Hill K.,
RA   Hitchcock P., Jackson P., Keim P., Longmire J., Lucas S., Okinaka R.,
RA   Richardson P., Rubin E., Tice H.;
RT   "Complete genome sequence of Bacillus cereus ZK.";
RL   Submitted (JUL-2004) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; CP000001; AAU20240.1; -.
KW   Kinase; Transferase.
SQ   SEQUENCE   523 AA;  59703 MW;  85B6AE9BB8CC2A3B CRC64;

  Query Match             39.7%;  Score 54;  DB 2;  Length 523;
  Best Local Similarity   54.5%;  Pred. No. 28;
  Matches   12;  Conservative    2;  Mismatches    8;  Indels    0;  Gaps    0;

Qy          3 VFSTEMANKAAEAVLKGQVETI 24
              :||      | |||||||| | :
Db        349 LFSLTFMTKGAEAVLKGQNEKV 370


RESULT 17
Q72X37
ID   Q72X37      PRELIMINARY;      PRT;  523 AA.
AC   Q72X37;
DT   05-JUL-2004 (TrEMBLrel. 27, Created)
DT   05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT   05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE   Sensor histidine kinase, putative (EC 2.7.3.-).
GN   OrderedLocusNames=BCE5541;
```

```
OS   Bacillus cereus (strain ATCC 10987).
OC   Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX   NCBI_TaxID=222523;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   PubMed=14960714; DOI=10.1093/nar/gkh258;
RA   Rasko D.A., Ravel J., Oekstad O.A., Helgason E., Cer R.Z., Jiang L.,
RA   Shores K.A., Fouts D.E., Tourasse N.J., Angiuoli S.V., Kolonay J.F.,
RA   Nelson W.C., Kolstoe A.-B., Fraser C.M., Read T.D.;
RT   "The genome sequence of Bacillus cereus ATCC 10987 reveals metabolic
RT   adaptations and a large plasmid related to Bacillus anthracis pXO1.";
RL   Nucleic Acids Res. 32:977-988(2004).
DR   EMBL; AE017281; AAS44441.1; -.
DR   TIGR; BCE5541; -.
DR   GO; GO:0005524; F:ATP binding; IEA.
DR   GO; GO:0016301; F:kinase activity; IEA.
DR   GO; GO:0016740; F:transferase activity; IEA.
DR   InterPro; IPR003594; ATPbind_ATPase.
DR   InterPro; IPR003018; GAF.
DR   Pfam; PF01590; GAF; 2.
DR   Pfam; PF02518; HATPase_c; 1.
DR   SMART; SM00065; GAF; 2.
KW   Complete proteome; Kinase; Transferase.
SQ   SEQUENCE   523 AA;  59731 MW;  4F105468CA527ABF CRC64;

  Query Match             39.7%;  Score 54;  DB 2;  Length 523;
  Best Local Similarity   54.5%;  Pred. No. 28;
  Matches   12;  Conservative    2;  Mismatches    8;  Indels    0;  Gaps    0;

Qy          3 VFSTEMANKAAEAVLKGQVETI 24
              :||      | ||||||| | :
Db        349 LFSLTFMTKGAEAVLKGQNEKV 370


RESULT 18
Q6HAM4
ID   Q6HAM4      PRELIMINARY;      PRT;   523 AA.
AC   Q6HAM4;
DT   05-JUL-2004 (TrEMBLrel. 27, Created)
DT   05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT   05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE   Two-component sensor protein.
GN   OrderedLocusNames=BT9727_5092;
OS   Bacillus thuringiensis (subsp. konkukian).
OC   Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX   NCBI_TaxID=180856;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=97-27;
RA   Brettin T.S., Bruce D., Challacombe J.F., Gilna P., Han C., Hill K.,
RA   Hitchcock P., Jackson P., Keim P., Longmire J., Lucas S., Okinaka R.,
RA   Richardson P., Rubin E., Tice H.;
RT   "Complete genome sequence of Bacillus thuringiensis 97-27.";
RL   Submitted (JUN-2004) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AE017355; AAT61192.1; -.
DR   GO; GO:0005524; F:ATP binding; IEA.
DR   InterPro; IPR003594; ATPbind_ATPase.
DR   InterPro; IPR003018; GAF.
DR   Pfam; PF01590; GAF; 2.
DR   Pfam; PF02518; HATPase_c; 1.
DR   SMART; SM00065; GAF; 2.
KW   Complete proteome.
SQ   SEQUENCE   523 AA;  59795 MW;  2C87843AEA3C3AA7 CRC64;

  Query Match             39.7%;  Score 54;  DB 2;  Length 523;
  Best Local Similarity   54.5%;  Pred. No. 28;
  Matches   12;  Conservative    2;  Mismatches    8;  Indels    0;  Gaps    0;

Qy          3 VFSTEMANKAAEAVLKGQVETI 24
              :||      | ||||||| | :
Db        349 LFSLTFMTKGAEAVLKGQNEKV 370


RESULT 19
Q71SH3
ID   Q71SH3      PRELIMINARY;      PRT;   516 AA.
AC   Q71SH3;
DT   05-JUL-2004 (TrEMBLrel. 27, Created)
DT   05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT   05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE   Putative aminotransferase.
OS   Oryza sativa (Rice).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC   Ehrhartoideae; Oryzeae; Oryza.
OX   NCBI_TaxID=4530;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Senescent leaf;
RA   Ansari M.I., Lee R.H., Chen S.C.G.;
RL   Submitted (AUG-2000) to the EMBL/GenBank/DDBJ databases.
CC   -!- SIMILARITY: Belongs to the class-III pyridoxal-phosphate-dependent
CC       aminotransferase family.
DR   EMBL; AF297651; AAQ14479.1; -.
DR   GO; GO:0030170; F:pyridoxal phosphate binding; IEA.
DR   GO; GO:0008483; F:transaminase activity; IEA.
DR   GO; GO:0016740; F:transferase activity; IEA.
DR   InterPro; IPR005814; Aminotrans_3.
DR   Pfam; PF00202; Aminotran_3; 1.
KW   Aminotransferase; Pyridoxal phosphate; Transferase.
SQ   SEQUENCE   516 AA;  56474 MW;  DCC7AC57563C403B CRC64;

  Query Match             39.0%;  Score 53;  DB 2;  Length 516;
  Best Local Similarity   45.8%;  Pred. No. 38;
  Matches   11;  Conservative    4;  Mismatches    9;  Indels    0;  Gaps    0;

Qy          4 FSTEMANKAAEAVLKGQVETIVSF 27
              |:| :||   | :||   ||| :|
Db        252 FATRLANNLEELILKEGPETIAAF 275


RESULT 20
Q7XN11
ID   Q7XN11      PRELIMINARY;      PRT;   516 AA.
AC   Q7XN11;
DT   01-OCT-2003 (TrEMBLrel. 25, Created)
DT   01-MAR-2004 (TrEMBLrel. 26, Last sequence update)
DT   01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE   OSJNBa0008M17.4 protein.
GN   Name=OSJNBa0008M17.4;
OS   Oryza sativa (japonica cultivar-group).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC   Ehrhartoideae; Oryzeae; Oryza.
OX   NCBI_TaxID=39947;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   PubMed=12447439; DOI=10.1038/nature01183;
RA   Feng Q., Zhang Y., Hao P., Wang S., Fu G., Huang Y., Li Y., Zhu J.,
RA   Liu Y., Hu X., Jia P., Zhang Y., Zhao Q., Ying K., Yu S., Tang Y.,
RA   Weng Q., Zhang L., Lu Y., Mu J., Lu Y., Zhang L.S., Yu Z., Fan D.,
RA   Liu X., Lu T., Li C., Wu Y., Sun T., Lei H., Li T., Hu H., Guan J.,
RA   Wu M., Zhang R., Zhou B., Chen Z., Chen L., Jin Z., Wang R., Yin H.,
RA   Cai Z., Ren S., Lv G., Gu W., Zhu G., Tu Y., Jia J., Zhang Y.,
RA   Chen J., Kang H., Chen X., Shao C., Sun Y., Hu Q., Zhang X., Zhang W.,
RA   Wang L., Ding C., Sheng H., Gu J., Chen S., Ni L., Zhu F., Chen W.,
RA   Lan L., Lai Y., Cheng Z., Gu M., Jiang J., Li J., Hong G., Xue Y.,
RA   Han B.;
RT   "Sequence and analysis of rice chromosome 4.";
RL   Nature 420:316-320(2002).
CC   -!- SIMILARITY: Belongs to the class-III pyridoxal-phosphate-dependent
CC       aminotransferase family.
DR   EMBL; AL662950; CAE04333.2; -.
DR   HSSP; P16932; 1DKA.
DR   Gramene; Q7XN11; -.
DR   GO; GO:0030170; F:pyridoxal phosphate binding; IEA.
```

```
DR   GO; GO:0008483; F:transaminase activity; IEA.
DR   InterPro; IPR005814; Aminotrans_3.
DR   Pfam; PF00202; Aminotran_3; 1.
KW   Pyridoxal phosphate.
SQ   SEQUENCE   516 AA;  56474 MW;  DCC7AC57563C403B CRC64;

  Query Match             39.0%;  Score 53;  DB 2;  Length 516;
  Best Local Similarity   45.8%;  Pred. No. 38;
  Matches   11;  Conservative    4;  Mismatches    9;  Indels    0;  Gaps    0;

Qy          4 FSTEMANKAAEAVLKGQVETIVSF 27
              |:| :||   | :||   ||| :|
Db        252 FATRLANNLEELILKEGPETIAAF 275


RESULT 21
Q6ZQL7
ID   Q6ZQL7      PRELIMINARY;      PRT;  1059 AA.
AC   Q6ZQL7;
DT   05-JUL-2004 (TrEMBLrel. 27, Created)
DT   05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT   05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE   Mus musculus cDNA fis, clone TRACH3031400, weakly similar to Valyl-
DE   tRNA synthetase 2 (EC 6.1.1.9).
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Kanehori K., Ishibashi T., Chiba Y., Fujimori K., Hiraoka S.,
RA   Tanai H., Watanabe S., Ishida S., Ono Y., Hotuta T., Watanabe M.,
RA   Sugiyama T., Irie R., Otsuki T., Sato H., Ota T., Wakamatsu A.,
RA   Ishii S., Yamamoto J., Isono Y., Kawai-Hio Y., Saito K., Nishikawa T.,
RA   Kimura K., Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Kanda K.,
RA   Wagatsuma M., Takahashi-Fujii A., Oshima A., Sugiyama A., Kawakami B.,
RA   Suzuki Y., Sugano S., Nagahari K., Masuho Y., Nagai K., Isogai T.;
RL   Submitted (JUL-2003) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AK129008; BAC87668.1; -.
DR   GO; GO:0005524; F:ATP binding; IEA.
DR   GO; GO:0004832; F:valine-tRNA ligase activity; IEA.
DR   GO; GO:0006438; P:valyl-tRNA aminoacylation; IEA.
DR   InterPro; IPR002300; tRNA-synt_1a.
DR   InterPro; IPR001412; tRNA-synt_I.
DR   InterPro; IPR002303; tRNA-synt_val.
DR   InterPro; IPR009080; tRNAsyn_1a_bind.
DR   InterPro; IPR009008; ValRS_IleRS_edit.
DR   Pfam; PF00133; tRNA-synt_1; 1.
DR   PRINTS; PR00986; TRNASYNTHVAL.
DR   TIGRFAMs; TIGR00422; valS; 1.
DR   PROSITE; PS00178; AA_TRNA_LIGASE_I; 1.
KW   Aminoacyl-tRNA synthetase.
SQ   SEQUENCE   1059 AA;  118369 MW;  0D617C315E8A7949 CRC64;

  Query Match             38.2%;  Score 52;  DB 2;  Length 1059;
  Best Local Similarity   44.4%;  Pred. No. 1.1e+02;
  Matches   12;  Conservative    5;  Mismatches   10;  Indels    0;  Gaps    0;

Qy          2 YVFSTEMANKAAEAVLKGQVETIVSFH 28
              :|   || ::||:||  | :|   |||
Db        481 FVRCQEMGDRAAKAVESGALELWPSFH 507


RESULT 22
Q6MG21
ID   Q6MG21      PRELIMINARY;      PRT;  1065 AA.
AC   Q6MG21;
DT   05-JUL-2004 (TrEMBLrel. 27, Created)
DT   05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT   05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE   KIAA1885 protein.
GN   Name=Kiaa1885;
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Brown Norway;
RX   PubMed=15060004; DOI=10.1101/gr.1987704;
RA   Hurt P., Walter L., Sudbrak R., Klages S., Mueller I., Shiina T.,
RA   Inoko H., Lehrach H., Guenther E., Reinhardt R., Himmelbauer H.;
RT   "The genomic sequence and comparative analysis of the rat major
RT   histocompatibility complex.";
RL   Genome Res. 14:631-639(2004).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Brown Norway;
RA   Boehm S., Borzym K., Gelling S., Gimmel V., Heitmann K., Kosiura A.,
RA   Lang N., Lehrack S., Thiel J., Sontag M., Hurt P., Himmelbauer H.,
RA   Sudbrak R., Reinhardt R.;
RL   Submitted (DEC-2003) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Brown Norway;
RA   Shiina T., Inoko H.;
RL   Submitted (DEC-2003) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BX883047; CAE84026.1; -.
DR   GO; GO:0005524; F:ATP binding; IEA.
DR   GO; GO:0004832; F:valine-tRNA ligase activity; IEA.
DR   GO; GO:0006438; P:valyl-tRNA aminoacylation; IEA.
DR   InterPro; IPR002300; tRNA-synt_1a.
DR   InterPro; IPR001412; tRNA-synt_I.
DR   InterPro; IPR002303; tRNA-synt_val.
DR   InterPro; IPR009080; tRNAsyn_1a_bind.
DR   InterPro; IPR009008; ValRS_IleRS_edit.
DR   Pfam; PF00133; tRNA-synt_1; 1.
DR   PRINTS; PR00986; TRNASYNTHVAL.
DR   TIGRFAMs; TIGR00422; valS; 1.
DR   PROSITE; PS00178; AA_TRNA_LIGASE_I; 1.
SQ   SEQUENCE   1065 AA;  118893 MW;  4BF57C6962DA50CC CRC64;

  Query Match             38.2%;  Score 52;  DB 2;  Length 1065;
  Best Local Similarity   44.4%;  Pred. No. 1.1e+02;
  Matches   12;  Conservative    5;  Mismatches   10;  Indels    0;  Gaps    0;

Qy          2 YVFSTEMANKAAEAVLKGQVETIVSFH 28
              :|   || ::||:||  | :|   |||
Db        481 FVRCQEMGDRAAKAVESGALELWPSFH 507


RESULT 23
Q69Z78
ID   Q69Z78      PRELIMINARY;      PRT;  1086 AA.
AC   Q69Z78;
DT   25-OCT-2004 (TrEMBLrel. 28, Created)
DT   25-OCT-2004 (TrEMBLrel. 28, Last sequence update)
DT   25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DE   MKIAA1885 protein (Fragment).
GN   Name=mKIAA1885;
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Embryonic intestinal tract;
RA   Okazaki N., Kikuno R.F., Ohara R., Inamoto S., Koseki H., Hiraoka S.,
RA   Saga Y., Seino S., Nishimura M., Kaisho T., Hoshino K., Kitamura H.,
RA   Nagase T., Ohara O., Koga H.;
RT   "Prediction of the Coding Sequences of Mouse Homologues of KIAA Gene:
RT   IV. The Complete Nucleotide Sequences of 500 Mouse KIAA-Homologous
RT   cDNAs Identified by Screening of Terminal Sequences of cDNA Clones
RT   Randomly Sampled from Size-Fractionated Libraries.";
RL   DNA Res. 11:205-218(2004).
DR   EMBL; AK173288; BAD32566.1; -.
```

```
DR   GO; GO:0005524; F:ATP binding; IEA.
DR   GO; GO:0004832; F:valine-tRNA ligase activity; IEA.
DR   GO; GO:0006438; P:valyl-tRNA aminoacylation; IEA.
DR   InterPro; IPR002300; tRNA-synt_1a.
DR   InterPro; IPR001412; tRNA-synt_I.
DR   InterPro; IPR002303; tRNA-synt_val.
DR   InterPro; IPR009080; tRNAsyn_1a_bind.
DR   Pfam; PF00133; tRNA-synt_1; 1.
DR   PRINTS; PR00986; TRNASYNTHVAL.
DR   TIGRFAMs; TIGR00422; valS; 1.
DR   PROSITE; PS00178; AA_TRNA_LIGASE_I; 1.
FT   NON_TER       1      1
SQ   SEQUENCE   1086 AA;  120787 MW;  DD957F2882106A62 CRC64;

  Query Match             38.2%;  Score 52;  DB 2;  Length 1086;
  Best Local Similarity   44.4%;  Pred. No. 1.2e+02;
  Matches   12;  Conservative    5;  Mismatches   10;  Indels    0;  Gaps    0;

Qy          2 YVFSTEMANKAAEAVLKGQVETIVSFH 28
              :|   || ::||:||  | :|   |||
Db        504 FVRCQEMGDRAAKAVESGALELWPSFH 530


RESULT 24
Q97JA0
ID   Q97JA0      PRELIMINARY;      PRT;   243 AA.
AC   Q97JA0;
DT   01-OCT-2001 (TrEMBLrel. 18, Created)
DT   01-OCT-2001 (TrEMBLrel. 18, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   Zn-dependent hydrolases, glyoxylase family.
GN   OrderedLocusNames=CAC1386;
OS   Clostridium acetobutylicum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1488;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 824 / DSM 792 / VKM B-1787;
RX   MEDLINE=21359325; PubMed=11466286;
RX   DOI=10.1128/JB.183.16.4823-4838.2001;
RA   Noelling J., Breton G., Omelchenko M.V., Makarova K.S., Zeng Q.,
RA   Gibson R., Lee H.M., Dubois J., Qiu D., Hitti J., Wolf Y.I.,
RA   Tatusov R.L., Sabathe F., Doucette-Stamm L.A., Soucaille P.,
RA   Daly M.J., Bennett G.N., Koonin E.V., Smith D.R.;
RT   "Genome sequence and comparative analysis of the solvent-producing
RT   bacterium Clostridium acetobutylicum.";
RL   J. Bacteriol. 183:4823-4838(2001).
DR   EMBL; AE007650; AAK79354.1; -.
DR   PIR; G97070; G97070.
DR   GO; GO:0016787; F:hydrolase activity; IEA.
DR   InterPro; IPR001279; Blactmase-like.
DR   Pfam; PF00753; Lactamase_B; 1.
KW   Complete proteome; Hydrolase.
SQ   SEQUENCE   243 AA;  26880 MW;  0F9F6A3EAADEBC0D CRC64;

  Query Match             37.5%;  Score 51;  DB 2;  Length 243;
  Best Local Similarity   25.0%;  Pred. No. 35;
  Matches    7;  Conservative   11;  Mismatches   10;  Indels    0;  Gaps    0;

Qy          1 VYVFSTEMANKAAEAVLKGQVETIVSFH 28
              :: | : :: |: | : |  :||:: :|
Db        199 LFDFDSNLSKKSLEKLTKYDIETVICYH 226


RESULT 25
Q84P52
ID   Q84P52      PRELIMINARY;      PRT;   520 AA.
AC   Q84P52;
DT   01-JUN-2003 (TrEMBLrel. 24, Created)
DT   01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT   01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE   Gamma-aminobutyrate transaminase subunit isozyme 3 (EC 2.6.1.19).
OS   Lycopersicon esculentum (Tomato).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids;
OC   lamiids; Solanales; Solanaceae; Solanum.
OX   NCBI_TaxID=4081;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Clark S.M., Shelp B.J.;
RL   Submitted (FEB-2003) to the EMBL/GenBank/DDBJ databases.
CC   -!- SIMILARITY: Belongs to the class-III pyridoxal-phosphate-dependent
CC       aminotransferase family.
DR   EMBL; AY240231; AAO92257.1; -.
DR   HSSP; P12995; 1QJ3.
DR   GO; GO:0003867; F:4-aminobutyrate transaminase activity; IEA.
DR   GO; GO:0030170; F:pyridoxal phosphate binding; IEA.
DR   GO; GO:0016740; F:transferase activity; IEA.
DR   InterPro; IPR005814; Aminotrans_3.
DR   Pfam; PF00202; Aminotran_3; 1.
DR   PROSITE; PS00600; AA_TRANSFER_CLASS_3; 1.
KW   Aminotransferase; Pyridoxal phosphate; Transferase.
SQ   SEQUENCE   520 AA;  57239 MW;  E4FCD1E922BD28F9 CRC64;

  Query Match             37.5%;  Score 51;  DB 2;  Length 520;
  Best Local Similarity   45.8%;  Pred. No. 77;
  Matches   11;  Conservative    3;  Mismatches   10;  Indels    0;  Gaps    0;

Qy          4 FSTEMANKAAEAVLKGQVETIVSF 27
              ||| :||     :||   ||| :|
Db        253 FSTRLANNLENLILKEGPETIAAF 276


RESULT 26
Q960E6
ID   Q960E6      PRELIMINARY;      PRT;  1049 AA.
AC   Q960E6;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE   SD04748p.
GN   Name=Aats-val;
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC   Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
RA   Champe M., Chavez C., Dorsett V., Farfan D., Frise E., George R.,
RA   Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Mungall C.J.,
RA   Nunoo J., Pacleb J., Paragas V., Park S., Phouanenavong S., Wan K.,
RA   Yu C., Lewis S.E., Rubin G.M., Celniker S.;
RL   Submitted (AUG-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AY052099; AAK93523.1; -.
DR   HSSP; P96142; 1IVS.
DR   FlyBase; FBgn0027079; Aats-val.
DR   GO; GO:0005524; F:ATP binding; IEA.
DR   GO; GO:0004832; F:valine-tRNA ligase activity; IEA.
DR   GO; GO:0006438; P:valyl-tRNA aminoacylation; IEA.
DR   InterPro; IPR002300; tRNA-synt_1a.
DR   InterPro; IPR001412; tRNA-synt_I.
DR   InterPro; IPR002303; tRNA-synt_val.
DR   InterPro; IPR009080; tRNAsyn_1a_bind.
DR   InterPro; IPR010978; tRNA_binding_arm.
DR   InterPro; IPR009008; ValRS_IleRS_edit.
DR   Pfam; PF00133; tRNA-synt_1; 1.
DR   PRINTS; PR00986; TRNASYNTHVAL.
DR   TIGRFAMs; TIGR00422; valS; 1.
DR   PROSITE; PS00178; AA_TRNA_LIGASE_I; 1.
SQ   SEQUENCE   1049 AA;  118331 MW;  56F322C7414EEAC4 CRC64;

  Query Match             37.5%;  Score 51;  DB 2;  Length 1049;
  Best Local Similarity   40.7%;  Pred. No. 1.6e+02;
```

```
  Matches   11;  Conservative    5;  Mismatches   11;  Indels    0;  Gaps    0;

Qy          2 YVFSTEMANKAAEAVLKGQVETIVSFH 28
              ||  ::||  | |||  |::: |   |
Db        471 YVSCSDMAASATEAVRSGELKIIPEHH 497


RESULT 27
Q9V6L1
ID   Q9V6L1      PRELIMINARY;      PRT;  1049 AA.
AC   Q9V6L1;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE   CG4062-PA (Cg4062-pb).
GN   Name=Aats-val; ORFNames=CG4062;
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC   Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20196006; PubMed=10731132; DOI=10.1126/science.287.5461.2185;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.H., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Gabor G.L.,
RA   Abril J.F., Agbayani A., An H.J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., WoodageT, Worley K.C., Wu D., Yang S., Yao Q.A., Ye J.,
RA   Yeh R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=22426065; PubMed=12537568;
RA   Celniker S.E., Wheeler D.A., Kronmiller B., Carlson J.W., Halpern A.,
RA   Patel S., Adams M., Champe M., Dugan S.P., Frise E., Hodgson A.,
RA   George R.A., Hoskins R.A., Laverty T., Muzny D.M., Nelson C.R.,
RA   Pacleb J.M., Park S., Pfeiffer B.D., Richards S., Sodergren E.J.,
RA   Svirskas R., Tabor P.E., Wan K., Stapleton M., Sutton G.G., Venter C.,
RA   Weinstock G., Scherer S.E., Myers E.W., Gibbs R.A., Rubin G.M.;
RT   "Finishing a whole-genome shotgun: Release 3 of the Drosophila
RT   melanogaster euchromatic genome sequence.";
RL   Genome Biol. 3:RESEARCH0079-RESEARCH0079(2002).
RN   [3]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=22426070; PubMed=12537573;
RA   Kaminker J.S., Bergman C.M., Kronmiller B., Carlson J., Svirskas R.,
RA   Patel S., Frise E., Wheeler D.A., Lewis S.E., Rubin G.M.,
RA   Ashburner M., Celniker S.E.;
RT   "The transposable elements of the Drosophila melanogaster euchromatin:
RT   a genomics perspective.";
RL   Genome Biol. 3:RESEARCH0084-RESEARCH0084(2002).
RN   [4]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=22426069; PubMed=12537572;
RA   Misra S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,
RA   Hradecky P., Huang Y., Kaminker J.S., Millburn G.H., Prochnik S.E.,
RA   Smith C.D., Tupy J.L., Whitfied E.J., Bayraktaroglu L., Berman B.P.,
RA   Bettencourt B.R., Celniker S.E., de Grey A.D., Drysdale R.A.,
RA   Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.Q.,
RA   Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,
RA   Lewis S.E.;
RT   "Annotation of the Drosophila melanogaster euchromatic genome: a
RT   systematic review.";
RL   Genome Biol. 3:RESEARCH0083-RESEARCH0083(2002).
RN   [5]
RP   SEQUENCE FROM N.A.
RG   FlyBase;
RL   Submitted (SEP-2002) to the EMBL/GenBank/DDBJ databases.
RN   [6]
RP   SEQUENCE FROM N.A.
RG   FlyBase;
RL   Submitted (MAR-2004) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AE003819; AAM68598.1; -.
DR   HSSP; P96142; 1IVS.
DR   IntAct; Q9V6L1; -.
DR   FlyBase; FBgn0027079; Aats-val.
DR   GO; GO:0005524; F:ATP binding; IEA.
DR   GO; GO:0004832; F:valine-tRNA ligase activity; IEA.
DR   GO; GO:0006438; P:valyl-tRNA aminoacylation; IEA.
DR   InterPro; IPR002300; tRNA-synt_1a.
DR   InterPro; IPR001412; tRNA-synt_I.
DR   InterPro; IPR002303; tRNA-synt_val.
DR   InterPro; IPR009080; tRNAsyn_1a_bind.
DR   InterPro; IPR010978; tRNA_binding_arm.
DR   InterPro; IPR009008; ValRS_IleRS_edit.
DR   Pfam; PF00133; tRNA-synt_1; 1.
DR   PRINTS; PR00986; TRNASYNTHVAL.
DR   TIGRFAMs; TIGR00422; valS; 1.
DR   PROSITE; PS00178; AA_TRNA_LIGASE_I; 1.
SQ   SEQUENCE   1049 AA;  118253 MW;  13A513ABF69E8EEB CRC64;

  Query Match             37.5%;  Score 51;  DB 2;  Length 1049;
  Best Local Similarity   40.7%;  Pred. No. 1.6e+02;
  Matches   11;  Conservative    5;  Mismatches   11;  Indels    0;  Gaps    0;

Qy          2 YVFSTEMANKAAEAVLKGQVETIVSFH 28
              ||  ::||  | |||  |::: |   |
Db        471 YVSCSDMAASATEAVRSGELKIIPEHH 497


RESULT 28
SYV_FUGRU
ID   SYV_FUGRU      STANDARD;      PRT;  1217 AA.
AC   P49696;
DT   01-FEB-1996 (Rel. 33, Created)
DT   01-FEB-1996 (Rel. 33, Last sequence update)
DT   25-OCT-2004 (Rel. 45, Last annotation update)
DE   Valyl-tRNA synthetase (EC 6.1.1.9) (Valine--tRNA ligase) (ValRS).
GN   Name=VARS1;
OS   Fugu rubripes (Japanese pufferfish) (Takifugu rubripes).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC   Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
OC   Tetradontoidea; Tetraodontidae; Takifugu.
```

```
OX   NCBI_TaxID=31033;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=97396021; PubMed=9254008;
RA   Lim E.H., Corrochano L.M., Elgar G., Brenner S.;
RT   "Genomic structure and sequence analysis of the valyl-tRNA synthetase
RT   gene of the Japanese pufferfish, Fugu rubripes.";
RL   DNA Seq. 7:141-151(1997).
CC   -!- CATALYTIC ACTIVITY: ATP + L-valine + tRNA(Val) = AMP + diphosphate
CC       + L-valyl-tRNA(Val).
CC   -!- SIMILARITY: Belongs to the class-I aminoacyl-tRNA synthetase
CC       family.
CC   -!- SIMILARITY: Contains 1 GST-like domain.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X91856; CAA62967.1; -.
DR   HSSP; P96142; 1IVS.
DR   InterPro; IPR010987; GST_C_like.
DR   InterPro; IPR004046; GST_Cterm.
DR   InterPro; IPR002300; tRNA-synt_1a.
DR   InterPro; IPR001412; tRNA-synt_I.
DR   InterPro; IPR002303; tRNA-synt_val.
DR   InterPro; IPR010978; tRNA_binding_arm.
DR   InterPro; IPR009080; tRNAsyn_1a_bind.
DR   InterPro; IPR009008; ValRS_IleRS_edit.
DR   Pfam; PF00043; GST_C; 1.
DR   Pfam; PF00133; tRNA-synt_1; 1.
DR   PRINTS; PR00986; TRNASYNTHVAL.
DR   TIGRFAMs; TIGR00422; valS; 1.
DR   PROSITE; PS00178; AA_TRNA_LIGASE_I; 1.
KW   Aminoacyl-tRNA synthetase; ATP-binding; Ligase; Protein biosynthesis.
FT   DOMAIN        1      ?       GST.
FT   SITE        293    303       "HIGH" region.
FT   SITE        809    813       "KMSKS" region.
FT   BINDING     812    812       ATP (By similarity).
SQ   SEQUENCE   1217 AA;  138218 MW;  5E08AF24B5C8A7A1 CRC64;

  Query Match             37.5%;  Score 51;  DB 1;  Length 1217;
  Best Local Similarity   37.0%;  Pred. No. 1.8e+02;
  Matches   10;  Conservative    8;  Mismatches    9;  Indels    0;  Gaps    0;

Qy          2 YVFSTEMANKAAEAVLKGQVETIVSFH 28
              ||  ::|  :||:|| :|::: |   |
Db        626 YVSCSDMGKQAADAVREGRLKIIPDHH 652


RESULT 29
Q89JQ9
ID   Q89JQ9      PRELIMINARY;      PRT;   338 AA.
AC   Q89JQ9;
DT   01-JUN-2003 (TrEMBLrel. 24, Created)
DT   01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT   01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE   Blr5213 protein.
GN   OrderedLocusNames=blr5213;
OS   Bradyrhizobium japonicum.
OC   Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC   Bradyrhizobiaceae; Bradyrhizobium.
OX   NCBI_TaxID=375;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=USDA110;
RX   MEDLINE=22484998; PubMed=12597275;
RA   Kaneko T., Nakamura Y., Sato S., Minamisawa K., Uchiumi T.,
RA   Sasamoto S., Watanabe A., Idesawa K., Iriguchi M., Kawashima K.,
RA   Kohara M., Matsumoto M., Shimpo S., Tsuruoka H., Wada T., Yamada M.,
RA   Tabata S.;
RT   "Complete genomic sequence of nitrogen-fixing symbiotic bacterium
RT   Bradyrhizobium japonicum USDA110.";
RL   DNA Res. 9:189-197(2002).
DR   EMBL; AP005954; BAC50478.1; -.
DR   GO; GO:0005524; F:ATP binding; IEA.
DR   GO; GO:0003910; F:DNA ligase (ATP) activity; IEA.
DR   GO; GO:0006310; P:DNA recombination; IEA.
DR   GO; GO:0006281; P:DNA repair; IEA.
DR   GO; GO:0006260; P:DNA replication; IEA.
DR   InterPro; IPR000977; DNA_ligase.
DR   Pfam; PF04679; DNA_ligase_A_C; 1.
DR   Pfam; PF01068; DNA_ligase_A_M; 1.
KW   Complete proteome.
SQ   SEQUENCE   338 AA;  37922 MW;  8E27957946E4E1D9 CRC64;

  Query Match             36.8%;  Score 50;  DB 2;  Length 338;
  Best Local Similarity   48.1%;  Pred. No. 70;
  Matches   13;  Conservative    3;  Mismatches   11;  Indels    0;  Gaps    0;

Qy          2 YVFSTEMANKAAEAVLKGQVETIVSFH 28
              :|  |     ||:| |  |:||||| |
Db        232 HVGFTSGLKSAAKAALTDQLETIVSDH 258


RESULT 30
O82521
ID   O82521      PRELIMINARY;      PRT;   459 AA.
AC   O82521;
DT   01-NOV-1998 (TrEMBLrel. 08, Created)
DT   01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT   01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE   Putative aminotransferase.
OS   Capsicum chinense (Scotch bonnet) (Bonnet pepper).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids;
OC   lamiids; Solanales; Solanaceae; Capsicum.
OX   NCBI_TaxID=80379;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=habanero;
RA   Aluru M., Curry J., O'Connell M.;
RT   "Nucleotide Sequence of a Probable Aminotransferase Gene (Accession
RT   No. AF085149) from Habanero Chile. (PGR98-182).";
RL   Plant Physiol. 118:1102-1102(1998).
CC   -!- SIMILARITY: Belongs to the class-III pyridoxal-phosphate-dependent
CC       aminotransferase family.
DR   EMBL; AF085149; AAC78480.1; -.
DR   HSSP; P04181; 2OAT.
DR   GO; GO:0030170; F:pyridoxal phosphate binding; IEA.
DR   GO; GO:0008483; F:transaminase activity; IEA.
DR   GO; GO:0016740; F:transferase activity; IEA.
DR   InterPro; IPR005814; Aminotrans_3.
DR   Pfam; PF00202; Aminotran_3; 1.
DR   PROSITE; PS00600; AA_TRANSFER_CLASS_3; 1.
KW   Aminotransferase; Pyridoxal phosphate; Transferase.
SQ   SEQUENCE   459 AA;  50729 MW;  02ABB4D728B524E4 CRC64;

  Query Match             36.8%;  Score 50;  DB 2;  Length 459;
  Best Local Similarity   41.7%;  Pred. No. 95;
  Matches   10;  Conservative    4;  Mismatches   10;  Indels    0;  Gaps    0;

Qy          4 FSTEMANKAAEAVLKGQVETIVSF 27
              ||| :||     :||   ||: :|
Db        196 FSTRLANNLESLILKEGPETVAAF 219


RESULT 31
Q814L6
ID   Q814L6      PRELIMINARY;      PRT;   523 AA.
AC   Q814L6;
DT   01-JUN-2003 (TrEMBLrel. 24, Created)
DT   01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
```

```
DT   01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE   Two-component sensor protein yhcY (EC 2.7.3.-).
GN   OrderedLocusNames=BC5412;
OS   Bacillus cereus (strain ATCC 14579 / DSM 31).
OC   Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX   NCBI_TaxID=226900;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=22608415; PubMed=12721630; DOI=10.1038/nature01582;
RA   Ivanova N., Sorokin A., Anderson I., Galleron N., Candelon B.,
RA   Kapatral V., Bhattacharyya A., Reznik G., Mikhailova N., Lapidus A.,
RA   Chu L., Mazur M., Goltsman E., Larsen N., D'Souza M., Walunas T.,
RA   Grechkin Y., Pusch G., Haselkorn R., Fonstein M., Ehrlich S.D.,
RA   Overbeek R., Kyrpides N.C.;
RT   "Genome sequence of Bacillus cereus and comparative analysis with
RT   Bacillus anthracis.";
RL   Nature 423:87-91(2003).
DR   EMBL; AE017015; AAP12274.1; -.
DR   GO; GO:0005524; F:ATP binding; IEA.
DR   GO; GO:0016740; F:transferase activity; IEA.
DR   InterPro; IPR003594; ATPbind_ATPase.
DR   InterPro; IPR003018; GAF.
DR   Pfam; PF01590; GAF; 2.
DR   Pfam; PF02518; HATPase_c; 1.
DR   SMART; SM00065; GAF; 2.
DR   SMART; SM00387; HATPase_c; 1.
KW   Complete proteome; Transferase.
SQ   SEQUENCE   523 AA;  59607 MW;  CAA8B8C8D0CE3E7D CRC64;

  Query Match             36.8%;  Score 50;  DB 2;  Length 523;
  Best Local Similarity   50.0%;  Pred. No. 1.1e+02;
  Matches   11;  Conservative    3;  Mismatches    8;  Indels    0;  Gaps    0;

Qy          3 VFSTEMANKAAEAVLKGQVETI 24
              :||      | |||||||: | :
Db        349 LFSLTFMTKGAEAVLKGKNEKV 370


RESULT 32
Q9H6R2
ID   Q9H6R2      PRELIMINARY;      PRT;   642 AA.
AC   Q9H6R2;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE   Hypothetical protein FLJ21965.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Kawabata A., Hikiji T., Kobatake N., Inagaki H., Ikema Y., Okamoto S.,
RA   Okitani R., Ota T., Suzuki Y., Obayashi M., Nishi T., Shibahara T.,
RA   Tanaka T., Nakamura Y., Isogai T., Sugano S.;
RL   Submitted (AUG-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AK025618; BAB15191.1; -.
DR   HSSP; P96142; 1IVS.
DR   GO; GO:0005524; F:ATP binding; IEA.
DR   GO; GO:0004832; F:valine-tRNA ligase activity; IEA.
DR   GO; GO:0006438; P:valyl-tRNA aminoacylation; IEA.
DR   InterPro; IPR002300; tRNA-synt_1a.
DR   InterPro; IPR002303; tRNA-synt_val.
DR   InterPro; IPR009080; tRNAsyn_1a_bind.
DR   InterPro; IPR009008; ValRS_IleRS_edit.
DR   Pfam; PF00133; tRNA-synt_1; 1.
DR   PRINTS; PR00986; TRNASYNTHVAL.
SQ   SEQUENCE   642 AA;  71578 MW;  C9E37EE1D742B7F1 CRC64;

  Query Match             36.8%;  Score 50;  DB 2;  Length 642;
  Best Local Similarity   44.4%;  Pred. No. 1.3e+02;
  Matches   12;  Conservative    4;  Mismatches   11;  Indels    0;  Gaps    0;

Qy          2 YVFSTEMANKAAEAVLKGQVETIVSFH 28
              :|   ||  :||:||  | :|   |||
Db         60 FVRCQEMGARAAKAVESGALELSPSFH 86


RESULT 33
Q6DKJ5
ID   Q6DKJ5      PRELIMINARY;      PRT;   657 AA.
AC   Q6DKJ5;
DT   25-OCT-2004 (TrEMBLrel. 28, Created)
DT   25-OCT-2004 (TrEMBLrel. 28, Last sequence update)
DT   25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DE   VARS2L protein (Fragment).
GN   Name=VARS2L;
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Brain;
RX   MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA   Krzywinski M.I., Skalska U., Smailus D.E., Schnerch A., Schein J.E.,
RA   Jones S.J., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length human
RT   and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Brain;
RA   Director MGC Project;
RL   Submitted (JUN-2004) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC073838; AAH73838.1; -.
DR   GO; GO:0005524; F:ATP binding; IEA.
DR   GO; GO:0004832; F:valine-tRNA ligase activity; IEA.
DR   GO; GO:0006438; P:valyl-tRNA aminoacylation; IEA.
DR   InterPro; IPR002300; tRNA-synt_1a.
DR   InterPro; IPR002303; tRNA-synt_val.
DR   InterPro; IPR009080; tRNAsyn_1a_bind.
DR   InterPro; IPR009008; ValRS_IleRS_edit.
DR   Pfam; PF00133; tRNA-synt_1; 1.
DR   PRINTS; PR00986; TRNASYNTHVAL.
FT   NON_TER       1      1
SQ   SEQUENCE   657 AA;  73196 MW;  BC34A3735FFA400A CRC64;

  Query Match             36.8%;  Score 50;  DB 2;  Length 657;
  Best Local Similarity   44.4%;  Pred. No. 1.4e+02;
  Matches   12;  Conservative    4;  Mismatches   11;  Indels    0;  Gaps    0;

Qy          2 YVFSTEMANKAAEAVLKGQVETIVSFH 28
              :|   ||  :||:||  | :|   |||
Db         75 FVRCQEMGARAAKAVESGALELSPSFH 101


RESULT 34
Q96GN2
ID   Q96GN2      PRELIMINARY;      PRT;   733 AA.
AC   Q96GN2;
```

```
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-MAR-2004 (TrEMBLrel. 26, Last sequence update)
DT   01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE   VARS2L protein (Fragment).
GN   Name=VARS2L;
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Skin;
RX   MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA   Krzywinski M.I., Skalska U., Smailus D.E., Schnerch A., Schein J.E.,
RA   Jones S.J., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length human
RT   and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Skin;
RA   Strausberg R.;
RL   Submitted (JUN-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC009355; AAH09355.2; -.
DR   HSSP; P96142; 1IVS.
DR   GO; GO:0005524; F:ATP binding; IEA.
DR   GO; GO:0004832; F:valine-tRNA ligase activity; IEA.
DR   GO; GO:0006438; P:valyl-tRNA aminoacylation; IEA.
DR   InterPro; IPR002110; ANK.
DR   InterPro; IPR002300; tRNA-synt_1a.
DR   InterPro; IPR002303; tRNA-synt_val.
DR   InterPro; IPR009080; tRNAsyn_1a_bind.
DR   InterPro; IPR009008; ValRS_IleRS_edit.
DR   Pfam; PF00133; tRNA-synt_1; 1.
DR   PRINTS; PR01415; ANKYRIN.
DR   PRINTS; PR00986; TRNASYNTHVAL.
FT   NON_TER       1      1
SQ   SEQUENCE   733 AA;  81230 MW;  B0433DC47AAB6721 CRC64;

  Query Match            36.8%;  Score 50;  DB 2;  Length 733;
  Best Local Similarity  44.4%;  Pred. No. 1.5e+02;
  Matches   12;  Conservative    4;  Mismatches   11;  Indels    0;  Gaps    0;

Qy          2 YVFSTEMANKAAEAVLKGQVETIVSFH 28
              :|   ||  :||:||  | :|   |||
Db        151 FVRCQEMGARAAKAVESGALELSPSFH 177


RESULT 35
Q96Q02
ID   Q96Q02      PRELIMINARY;      PRT;  1098 AA.
AC   Q96Q02;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE   KIAA1885 protein (Fragment).
GN   Name=KIAA1885;
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Brain;
RX   MEDLINE=21456161; PubMed=11572484;
RA   Nagase T., Kikuno R., Ohara O.;
RT   "Prediction of the coding sequences of unidentified human genes. XXI.
RT   The complete sequences of 60 new cDNA clones from brain which code for
RT   large proteins.";
RL   DNA Res. 8:179-187(2001).
DR   EMBL; AB067472; BAB67778.1; -.
DR   HSSP; P96142; 1IVS.
DR   Genew; HGNC:21642; VARS2L.
DR   GO; GO:0005524; F:ATP binding; IEA.
DR   GO; GO:0004832; F:valine-tRNA ligase activity; IEA.
DR   GO; GO:0006438; P:valyl-tRNA aminoacylation; IEA.
DR   InterPro; IPR002300; tRNA-synt_1a.
DR   InterPro; IPR001412; tRNA-synt_I.
DR   InterPro; IPR002303; tRNA-synt_val.
DR   InterPro; IPR009080; tRNAsyn_1a_bind.
DR   InterPro; IPR009008; ValRS_IleRS_edit.
DR   Pfam; PF00133; tRNA-synt_1; 1.
DR   PRINTS; PR00986; TRNASYNTHVAL.
DR   TIGRFAMs; TIGR00422; valS; 1.
DR   PROSITE; PS00178; AA_TRNA_LIGASE_I; 1.
FT   NON_TER       1      1
SQ   SEQUENCE   1098 AA;  122469 MW;  E01DCA8C8E42BC4D CRC64;

  Query Match            36.8%;  Score 50;  DB 2;  Length 1098;
  Best Local Similarity  44.4%;  Pred. No. 2.3e+02;
  Matches   12;  Conservative    4;  Mismatches   11;  Indels    0;  Gaps    0;

Qy          2 YVFSTEMANKAAEAVLKGQVETIVSFH 28
              :|   ||  :||:||  | :|   |||
Db        516 FVRCQEMGARAAKAVESGALELSPSFH 542


RESULT 36
Q8GGI9
ID   Q8GGI9      PRELIMINARY;      PRT;   350 AA.
AC   Q8GGI9;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE   Abc1.
GN   Name=abc1;
OS   Lactobacillus plantarum.
OC   Bacteria; Firmicutes; Lactobacillales; Lactobacillaceae;
OC   Lactobacillus.
OX   NCBI_TaxID=1590;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CCM3626;
RA   Bringel F., Hubert J.-C.;
RL   Submitted (MAY-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF514870; AAO15990.1; -.
SQ   SEQUENCE   350 AA;  38178 MW;  64D1986ED73C8AAA CRC64;

  Query Match            36.4%;  Score 49.5;  DB 2;  Length 350;
  Best Local Similarity  54.5%;  Pred. No. 86;
  Matches   12;  Conservative    4;  Mismatches    3;  Indels    3;  Gaps    1;

Qy          3 VFSTEMANKAAEAVLKGQVETI 24
              |:||::  ||||   ||||: |
Db        176 VYSTDLLAKAAE---KGQVDAI 194


RESULT 37
Q9LFW7
ID   Q9LFW7      PRELIMINARY;      PRT;   141 AA.
AC   Q9LFW7;
```

```
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   T7N9.31.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Khan S., Brooks S., Buehler E., Chao Q., Johnson-Hopson C., Kim C.,
RA   Shinn P., Altafi H., Bei Q., Chin C., Chiou J., Choi E., Conn L.,
RA   Conway A., Gonzales A., Hansen N., Howng B., Koo T., Lam B., Lee J.,
RA   Lenz C., Li J., Liu A., Liu K., Liu S., Mukharsky N., Nguyen M.,
RA   Palm C., Pham P., Sakano H., Schwartz J., Southwick A., Thaveri A.,
RA   Toriumi M., Vaysberg M., Yu G., Federspiel N.A., Theologis A.,
RA   Ecker J.R.;
RL   Submitted (MAY-2000) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Cheuk R., Shinn P., Brooks S., Buehler E., Chao Q., Johnson-Hopson C.,
RA   Khan S., Kim C., Altafi H., Bei B., Chin C., Chiou J., Choi E.,
RA   Conn L., Conway A., Gonzalez A., Hansen N., Howing B., Koo T., Lam B.,
RA   Lee J., Lenz C., Li J., Liu A., Liu J., Liu S., Mukharsky N.,
RA   Nguyen M., Palm C., Pham P., Sakano H., Schwartz J., Southwick A.,
RA   Thaveri A., Toriumi M., Vaysberg M., Yu G., Davis R., Federspiel N.,
RA   Theologis A., Ecker J.;
RL   Submitted (JUN-2000) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RA   Cheuk R., Shinn P., Brooks S., Buehler E., Chao Q., Johnson-Hopson C.,
RA   Khan S., Kim C., Altafi H., Bei B., Chin C., Chiou J., Choi E.,
RA   Conn L., Conway A., Gonzalez A., Hansen N., Howing B., Koo T., Lam B.,
RA   Lee J., Lenz C., Li J., Liu A., Liu J., Liu S., Mukharsky N.,
RA   Nguyen M., Palm C., Pham P., Sakano H., Schwartz J., Southwick A.,
RA   Thaveri A., Toriumi M., Vaysberg M., Yu G., Davis R., Federspiel N.,
RA   Theologis A., Ecker J.;
RL   Submitted (SEP-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AC000348; AAF79866.1; -.
SQ   SEQUENCE   141 AA;  16649 MW;  6FEC48B07326B313 CRC64;

  Query Match             36.0%;  Score 49;  DB 2;  Length 141;
  Best Local Similarity   41.7%;  Pred. No. 40;
  Matches   10;  Conservative    6;  Mismatches    8;  Indels    0;  Gaps    0;

Qy          4 FSTEMANKAAEAVLKGQVETIVSF 27
              :|    ||||: : ||::|  |:|
Db        112 WSFRSTNKAADRLAKGELENNVTF 135


RESULT 38
Q976C9
ID   Q976C9      PRELIMINARY;      PRT;   143 AA.
AC   Q976C9;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   Hypothetical protein ST0252.
GN   OrderedLocusNames=ST0252;
OS   Sulfolobus tokodaii.
OC   Archaea; Crenarchaeota; Thermoprotei; Sulfolobales; Sulfolobaceae;
OC   Sulfolobus.
OX   NCBI_TaxID=111955;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=JCM 10545 / 7;
RX   MEDLINE=21456156; PubMed=11572479;
RA   Kawarabayasi Y., Hino Y., Horikawa H., Jin-no K., Takahashi M.,
RA   Sekine M., Baba S.-I., Ankai A., Kosugi H., Hosoyama A., Fukui S.,
RA   Nagai Y., Nishijima K., Otsuka R., Nakazawa H., Takamiya M., Kato Y.,
RA   Yoshizawa T., Tanaka T., Kudoh Y., Yamazaki J., Kushida N., Oguchi A.,
RA   Aoki K.-I., Masuda S., Yanagii M., Nishimura M., Yamagishi A.,
RA   Oshima T., Kikuchi H.;
RT   "Complete genome sequence of an aerobic thermoacidophilic
RT   Crenarchaeon, Sulfolobus tokodaii strain7.";
RL   DNA Res. 8:123-140(2001).
DR   EMBL; AP000982; BAB65218.1; -.
KW   Complete proteome; Hypothetical protein.
SQ   SEQUENCE   143 AA;  16822 MW;  D1AFCC5157298CA2 CRC64;

  Query Match             36.0%;  Score 49;  DB 2;  Length 143;
  Best Local Similarity   43.3%;  Pred. No. 41;
  Matches   13;  Conservative    4;  Mismatches   11;  Indels    2;  Gaps    1;

Qy          1 VYVFSTEMAN--KAAEAVLKGQVETIVSFH 28
              | |  | | |  | || ||| ::: :|  |
Db         26 VLVVLTNMKNVEKEAEKVLKTRIDKVVYIH 55


RESULT 39
Q8G3S3
ID   Q8G3S3      PRELIMINARY;      PRT;   187 AA.
AC   Q8G3S3;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE   Hypoxanthine-guanine phosphoribosyltransferase.
GN   Name=hprT; OrderedLocusNames=BL1681;
OS   Bifidobacterium longum.
OC   Bacteria; Actinobacteria; Actinobacteridae; Bifidobacteriales;
OC   Bifidobacteriaceae; Bifidobacterium.
OX   NCBI_TaxID=216816;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=NCC 2705;
RX   MEDLINE=22294977; PubMed=12381787; DOI=10.1073/pnas.212527599;
RA   Schell M.A., Karmirantzou M., Snel B., Vilanova D., Berger B.,
RA   Pessi G., Zwahlen M.-C., Desiere F., Bork P., Delley M.,
RA   Pridmore R.D., Arigoni F.;
RT   "The genome sequence of Bifidobacterium longum reflects its adaptation
RT   to the human gastrointestinal tract.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:14422-14427(2002).
CC   -!- SIMILARITY: Belongs to the purine/pyrimidine
CC       phosphoribosyltransferase family.
DR   EMBL; AE014802; AAN25468.1; -.
DR   HSSP; O33799; 1J7J.
DR   GO; GO:0005737; C:cytoplasm; IEA.
DR   GO; GO:0004422; F:hypoxanthine phosphoribosyltransferase acti. . .; IEA.
DR   GO; GO:0016740; F:transferase activity; IEA.
DR   GO; GO:0009116; P:nucleoside metabolism; IEA.
DR   GO; GO:0006166; P:purine ribonucleoside salvage; IEA.
DR   InterPro; IPR002110; ANK.
DR   InterPro; IPR005904; Hxn_phspho_trans.
DR   InterPro; IPR002375; Pr/py_rp_transf.
DR   InterPro; IPR000836; PRtransferase.
DR   Pfam; PF00156; Pribosyltran; 1.
DR   PRINTS; PR01415; ANKYRIN.
DR   TIGRFAMs; TIGR01203; HGPRTase; 1.
DR   PROSITE; PS00103; PUR_PYR_PR_TRANSFER; 1.
KW   Complete proteome; Glycosyltransferase; Transferase.
SQ   SEQUENCE   187 AA;  20617 MW;  E697C3C127277DB1 CRC64;

  Query Match             36.0%;  Score 49;  DB 2;  Length 187;
  Best Local Similarity   45.2%;  Pred. No. 54;
  Matches   14;  Conservative    3;  Mismatches    4;  Indels   10;  Gaps    1;

Qy          7 EMANKAAE----------AVLKGQVETIVSF 27
              |||  |:|          ||||| | |:|:|
Db         26 EMAALASEDYRDKNPLLVAVLKGAVNTLVAF 56


RESULT 40
Q9FZL1
ID   Q9FZL1      PRELIMINARY;      PRT;   210 AA.
AC   Q9FZL1;
```

```
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   F17L21.2.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Johnson-Hopson C., Brooks S., Buehler E., Chao Q., Khan S., Kim C.,
RA   Shinn P., Altafi H., Bei Q., Chin C., Chiou J., Choi E., Conn L.,
RA   Conway A., Gonzales A., Hansen N., Howing B., Koo T., Lam B., Lee J.,
RA   Lenz C., Li J., Liu A., Liu K., Liu S., Mukharsky N., Nguyen M.,
RA   Palm C., Pham P., Sakano H., Schwartz J., Southwick A., Thaveri A.,
RA   Toriumi M., Vaysberg M., Yu G., Federspiel N.A., Theologis A.,
RA   Ecker J.R.;
RL   Submitted (AUG-2000) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Cheuk R., Shinn P., Brooks S., Buehler E., Chao Q., Johnson-Hopson C.,
RA   Khan S., Kim C., Altafi H., Bei B., Chin C., Chiou J., Choi E.,
RA   Conn L., Conway A., Gonzalez A., Hansen N., Howing B., Koo T., Lam B.,
RA   Lee J., Lenz C., Li J., Liu A., Liu J., Liu S., Mukharsky N.,
RA   Nguyen M., Palm C., Pham P., Sakano H., Schwartz J., Southwick A.,
RA   Thaveri A., Toriumi M., Vaysberg M., Yu G., Davis R., Federspiel N.,
RA   Theologis A., Ecker J.;
RL   Submitted (AUG-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AC004557; AAF99722.1; -.
DR   PIR; D86398; D86398.
DR   GO; GO:0005634; C:nucleus; IEA.
DR   GO; GO:0004089; F:carbonate dehydratase activity; IEA.
DR   GO; GO:0008270; F:zinc ion binding; IEA.
DR   GO; GO:0015976; P:carbon utilization; IEA.
DR   GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR   InterPro; IPR003822; PAH.
DR   InterPro; IPR001765; Prok_plnt_COanhd.
DR   Pfam; PF02671; PAH; 1.
SQ   SEQUENCE   210 AA;  24697 MW;  FCD8130CD75700A0 CRC64;

  Query Match             36.0%;  Score 49;  DB 2;  Length 210;
  Best Local Similarity   41.7%;  Pred. No. 60;
  Matches   10;  Conservative    6;  Mismatches    8;  Indels    0;  Gaps    0;

Qy          4 FSTEMANKAAEAVLKGQVETIVSF 27
              :|    ||||: : ||::|  |:|
Db        181 WSFRSTNKAADRLAKGELENNVTF 204


Search completed: June  8, 2005, 03:22:56
Job time : 136.25 secs
```

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: June 8, 2005, 03:00:28 ; Search time 164.25 Seconds
(without alignments)
112.237 Million cell updates/sec

Title: US-09-915-543-15_COPY_349_384
Perfect score: 183
Sequence: 1 DGLSQEQLEHRERSLQTLRDIQRMLFPDEKEFTGAQ 36

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1612378 seqs, 512079187 residues

Total number of hits satisfying chosen parameters: 1612378

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : UniProt_03:*
1: uniprot_sprot:*
2: uniprot_trembl:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	183	100.0	1426	1	BCL9_HUMAN	O00512 homo sapien
2	173	94.5	1425	2	Q67FX9	Q67fx9 mus musculu
3	163	89.1	796	2	Q6NRE2	Q6nre2 xenopus lae
4	147	80.3	1474	2	Q67FY0	Q67fy0 brachydanio
5	113	61.7	1530	2	Q67FY3	Q67fy3 brachydanio
6	109	59.6	1457	2	Q641L9	Q64119 mus musculu
7	109	59.6	1494	2	Q67FY1	Q67fy1 homo sapien
8	109	59.6	1494	2	Q67FY2	Q67fy2 mus musculu
9	109	59.6	1494	2	Q6I7B5	Q6i7b5 mus musculu
10	109	59.6	1499	2	Q86UU0	Q86uu0 homo sapien
11	66	36.1	1469	1	BCL9_DROME	Q961d9 drosophila
12	62	33.9	192	2	Q7RLG2	Q7rlg2 plasmodium
13	61	33.3	833	2	Q89YQ1	Q89yq1 bacteroides
14	58	31.7	102	2	Q7VNS5	Q7vns5 haemophilus
15	57	31.1	1034	2	Q8Z8Y6	Q8z8y6 salmonella
16	57	31.1	1036	2	Q83SZ9	Q83sz9 salmonella
17	57	31.1	1046	2	Q8ZRE3	Q8zre3 salmonella
18	56.5	30.9	476	2	Q7Q3P6	Q7q3p6 anopheles g
19	56	30.6	584	1	TRF4_YEAST	P53632 saccharomyc
20	56	30.6	818	2	Q8TGZ1	Q8tgz1 methanopyru
21	56	30.6	859	2	Q9NT51	Q9nt51 homo sapien
22	56	30.6	1150	2	Q9Y2H2	Q9y2h2 homo sapien
23	56	30.6	1208	2	Q9VXU1	Q9vxu1 drosophila
24	56	30.6	1398	2	Q9VXU2	Q9vxu2 drosophila
25	55.5	30.3	767	2	Q81U73	Q81u73 bacillus an
26	55.5	30.3	974	2	Q63EZ1	Q63ez1 bacillus ce
27	55	30.1	287	2	Q8UCA0	Q8uca0 agrobacteri
28	55	30.1	295	2	Q7CWN9	Q7cwn9 agrobacteri
29	55	30.1	335	2	Q6PCC7	Q6pcc7 homo sapien
30	55	30.1	337	2	Q6NVX9	Q6nvx9 homo sapien
31	55	30.1	426	1	HEM1_SYNEL	Q8di53 synechococc
32	55	30.1	428	2	Q6PJW1	Q6pjw1 homo sapien
33	55	30.1	433	2	Q8IZM6	Q8izm6 homo sapien
34	55	30.1	464	2	Q86XV6	Q86xv6 homo sapien
35	55	30.1	718	1	RHG8_HUMAN	Q9nsg0 homo sapien
36	54.5	29.8	174	1	ASC3_MOUSE	Q9jjr7 mus musculu
37	54.5	29.8	1171	1	PHYB_ORYSA	P25764 oryza sativ
38	54.5	29.8	1171	2	Q84LN8	Q84ln8 oryza sativ
39	54	29.5	326	2	Q6D1R6	Q6d1r6 erwinia car
40	54	29.5	554	2	Q7R1P0	Q7r1p0 giardia lam
41	54	29.5	832	2	Q64WV9	Q64wv9 bacteroides
42	53.5	29.2	155	2	Q8VD56	Q8vd56 rattus norv
43	53.5	29.2	208	2	Q6PZ60	Q6pz60 mycobacteri
44	53.5	29.2	266	2	Q8Y9X2	Q8y9x2 listeria mo
45	53.5	29.2	974	2	Q73CF5	Q73cf5 bacillus ce

ALIGNMENTS

RESULT 1
BCL9_HUMAN
ID BCL9_HUMAN STANDARD; PRT; 1426 AA.
AC O00512;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 05-JUL-2004 (Rel. 44, Last annotation update)
DE B-cell lymphoma 9 protein (Bcl-9) (Legless homolog).
GN Name=BCL9;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Fetal brain;
RX MEDLINE=98158621; PubMed=9490669;
RA Willis T.G., Zalcberg I.R., Coignet L.J.A., Wlodarska I., Stul M.,
RA Jadayel D.M., Bastard C., Treleaven J.G., Catovsky D., Silva M.L.M.,
RA Dyer M.J.S.;
RT "Molecular cloning of translocation t(1;14)(q21;q32) defines a novel
RT gene (BCL9) at chromosome 1q21.";
RL Blood 91:1873-1881(1998).
RN [2]
RP FUNCTION.
RX MEDLINE=21952490; PubMed=11955446; DOI=10.1016/S0092-8674(02)00679-7;
RA Kramps T., Peter O., Brunner E., Nellen D., Froesch B., Chatterjee S.,
RA Murone M., Zuellig S., Basler K.;
RT "Wnt/wingless signaling requires BCL9/legless-mediated recruitment of
RT pygopus to the nuclear beta-catenin-TCF complex.";
RL Cell 109:47-60(2002).
CC -!- FUNCTION: Involved in signal transduction through the Wnt pathway.
CC -!- SUBUNIT: Binds to beta-catenin (CTNNB1), PYGO1 and PYGO2.
CC -!- SUBCELLULAR LOCATION: Nuclear (Probable).
CC -!- TISSUE SPECIFICITY: Detected at low levels in thymus, prostate,
CC testis, ovary and small intestine, and at lower levels in spleen,
CC colon and blood.
CC -!- DISEASE: Involved in a t(1;14)(q21;q32) chromosomal translocation
CC found in a patient with precusor B-cell acute lymphoblastic
CC leukemia (ALL). This translocation leaves the coding region
CC intact, but may have pathogenic effects due to alterations in the
CC expression level of BCL9. Several cases of translocations within
CC the 3' untranslated region of BCL9 have been found in B-cell
CC malignancies.
CC -!- CAUTION: It is uncertain whether Met-1 or Met-27 is the initiator.
CC -!- CAUTION: Ref.1 sequence differs from that shown due to a
CC frameshift in position 1391.
CC ---

```
CC   or send an email to license@isb-sib.ch).
CC   ------------------------------------------------------------------------
DR   EMBL; Y13620; CAA73942.1; ALT_FRAME.
DR   Genew; HGNC:1008; BCL9.
DR   MIM; 602597; -.
KW   Chromosomal translocation; Nuclear protein; Proto-oncogene;
KW   Wnt signaling pathway.
FT   DOMAIN       231   1378       Pro-rich.
FT   DOMAIN       347    377       CTNNB1-binding.
FT   DOMAIN       331    335       Poly-Pro.
FT   DOMAIN       514    517       Poly-Pro.
FT   DOMAIN       900    903       Poly-Ala.
FT   DOMAIN       970    973       Poly-Pro.
SQ   SEQUENCE   1426 AA;  149314 MW;  A240A487716B7F1B CRC64;

  Query Match             100.0%;  Score 183;  DB 1;  Length 1426;
  Best Local Similarity   100.0%;  Pred. No. 2.7e-14;
  Matches   36;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 DGLSQEQLEHRERSLQTLRDIQRMLFPDEKEFTGAQ 36
              |||||||||||||||||||||||||||||||||||||
Db        349 DGLSQEQLEHRERSLQTLRDIQRMLFPDEKEFTGAQ 384


RESULT 2
Q67FX9
ID   Q67FX9      PRELIMINARY;      PRT;  1425 AA.
AC   Q67FX9;
DT   25-OCT-2004 (TrEMBLrel. 28, Created)
DT   25-OCT-2004 (TrEMBLrel. 28, Last sequence update)
DT   25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DE   BCL9.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6;
RA   Brembeck F.H., Schwarz-Romond T., Bakkers J., Wilhelm S.,
RA   Hammerschmidt M., Birchmeier W.;
RT   "Essential role of BCL9-2 in the switch between [beta]-catenin's
RT   adhesive and transcriptional functions.";
RL   Genes Dev. 18:0-0(2004).
DR   EMBL; AY296061; AAQ62699.1; -.
SQ   SEQUENCE   1425 AA;  148970 MW;  77347CF56FC4A815 CRC64;

  Query Match             94.5%;  Score 173;  DB 2;  Length 1425;
  Best Local Similarity   94.4%;  Pred. No. 5.1e-13;
  Matches   34;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 DGLSQEQLEHRERSLQTLRDIQRMLFPDEKEFTGAQ 36
              |||||||||||||||||||||||||||||||||  |
Db        349 DGLSQEQLEHRERSLQTLRDIQRMLFPDEKEFTAGQ 384


RESULT 3
Q6NRE2
ID   Q6NRE2      PRELIMINARY;      PRT;   796 AA.
AC   Q6NRE2;
DT   05-JUL-2004 (TrEMBLrel. 27, Created)
DT   05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT   05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE   MGC83888 protein.
GN   Name=MGC83888;
OS   Xenopus laevis (African clawed frog).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
OC   Xenopodinae; Xenopus.
OX   NCBI_TaxID=8355;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Oocytes;
RX   MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA   Krzywinski M.I., Skalska U., Smailus D.E., Schnerch A., Schein J.E.,
RA   Jones S.J., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length human
RT   and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Oocytes;
RX   MEDLINE=22341132; PubMed=12454917; DOI=10.1002/dvdy.10174;
RA   Klein S.L., Strausberg R.L., Wagner L., Pontius J., Clifton S.W.,
RA   Richardson P.;
RT   "Genetic and genomic tools for Xenopus research: The NIH Xenopus
RT   initiative.";
RL   Dev. Dyn. 225:384-391(2002).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Oocytes;
RA   Klein S., Strausberg R.;
RL   Submitted (MAY-2004) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC070813; AAH70813.1; -.
SQ   SEQUENCE   796 AA;  86048 MW;  9A282C1DCA316678 CRC64;

  Query Match             89.1%;  Score 163;  DB 2;  Length 796;
  Best Local Similarity   100.0%;  Pred. No. 5e-12;
  Matches   32;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 DGLSQEQLEHRERSLQTLRDIQRMLFPDEKEF 32
              ||||||||||||||||||||||||||||||||
Db        343 DGLSQEQLEHRERSLQTLRDIQRMLFPDEKEF 374


RESULT 4
Q67FY0
ID   Q67FY0      PRELIMINARY;      PRT;  1474 AA.
AC   Q67FY0;
DT   25-OCT-2004 (TrEMBLrel. 28, Created)
DT   25-OCT-2004 (TrEMBLrel. 28, Last sequence update)
DT   25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DE   Bcl9.
OS   Brachydanio rerio (Zebrafish) (Danio rerio).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC   Cyprinidae; Danio.
OX   NCBI_TaxID=7955;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Brembeck F.H., Schwarz-Romond T., Bakkers J., Wilhelm S.,
RA   Hammerschmidt M., Birchmeier W.;
RT   "Essential role of BCL9-2 in the switch between [beta]-catenin's
RT   adhesive and transcriptional functions.";
RL   Genes Dev. 18:0-0(2004).
DR   EMBL; AY296060; AAQ62698.1; -.
SQ   SEQUENCE   1474 AA;  154339 MW;  4B2C3E8092BE3532 CRC64;

  Query Match             80.3%;  Score 147;  DB 2;  Length 1474;
  Best Local Similarity   90.3%;  Pred. No. 1.1e-09;
  Matches   28;  Conservative    3;  Mismatches    0;  Indels    0;  Gaps    0;
```

```
Qy          1 DGLSQEQLEHRERSLQTLRDIQRMLFPDEKE 31
              :|||||||||||||||||||||||||||||:|:
Db        403 EGLSQEQLEHRERSLQTLRDIQRMLFPDDKD 433


RESULT 5
Q67FY3
ID   Q67FY3      PRELIMINARY;      PRT;  1530 AA.
AC   Q67FY3;
DT   25-OCT-2004 (TrEMBLrel. 28, Created)
DT   25-OCT-2004 (TrEMBLrel. 28, Last sequence update)
DT   25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DE   Bcl9-2.
OS   Brachydanio rerio (Zebrafish) (Danio rerio).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC   Cyprinidae; Danio.
OX   NCBI_TaxID=7955;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Brembeck F.H., Schwarz-Romond T., Bakkers J., Wilhelm S.,
RA   Hammerschmidt M., Birchmeier W.;
RT   "Essential role of BCL9-2 in the switch between [beta]-catenin's
RT   adhesive and transcriptional functions.";
RL   Genes Dev. 18:0-0(2004).
DR   EMBL; AY296057; AAQ62695.1; -.
SQ   SEQUENCE   1530 AA;  159872 MW;  C29FEC9433ED28C0 CRC64;

  Query Match             61.7%;  Score 113;  DB 2;  Length 1530;
  Best Local Similarity   88.0%;  Pred. No. 2.4e-05;
  Matches   22;  Conservative    3;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 DGLSQEQLEHRERSLQTLRDIQRML 25
              ||||:|||||||||||||||:|:|
Db        371 DGLSKEQLEHRERSLQTLRDIERLL 395


RESULT 6
Q641L9
ID   Q641L9      PRELIMINARY;      PRT;  1457 AA.
AC   Q641L9;
DT   25-OCT-2004 (TrEMBLrel. 28, Created)
DT   25-OCT-2004 (TrEMBLrel. 28, Last sequence update)
DT   25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DE   Bcl9l protein.
GN   Name=Bcl9l;
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6; TISSUE=Brain;
RX   PubMed=12477932; DOI=10.1073/pnas.242603899;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA   Krzywinski M.I., Skalska U., Smailus D.E., Schnerch A., Schein J.E.,
RA   Jones S.J., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length human
RT   and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6; TISSUE=Brain;
RA   Director MGC Project;
RL   Submitted (SEP-2004) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC082304; AAH82304.1; -.
SQ   SEQUENCE   1457 AA;  152636 MW;  4FD2B47ADDE92A33 CRC64;

  Query Match             59.6%;  Score 109;  DB 2;  Length 1457;
  Best Local Similarity   84.0%;  Pred. No. 7.3e-05;
  Matches   21;  Conservative    4;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 DGLSQEQLEHRERSLQTLRDIQRML 25
              :|||:||||||||||||||||:|:|
Db        357 EGLSKEQLEHRERSLQTLRDIERLL 381


RESULT 7
Q67FY1
ID   Q67FY1      PRELIMINARY;      PRT;  1494 AA.
AC   Q67FY1;
DT   25-OCT-2004 (TrEMBLrel. 28, Created)
DT   25-OCT-2004 (TrEMBLrel. 28, Last sequence update)
DT   25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DE   BCL9-2.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Brembeck F.H., Schwarz-Romond T., Bakkers J., Wilhelm S.,
RA   Hammerschmidt M., Birchmeier W.;
RT   "Essential role of BCL9-2 in the switch between [beta]-catenin's
RT   adhesive and transcriptional functions.";
RL   Genes Dev. 18:0-0(2004).
DR   EMBL; AY296059; AAQ62697.1; -.
SQ   SEQUENCE   1494 AA;  156528 MW;  2D591F45FF3AEE36 CRC64;

  Query Match             59.6%;  Score 109;  DB 2;  Length 1494;
  Best Local Similarity   84.0%;  Pred. No. 7.5e-05;
  Matches   21;  Conservative    4;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 DGLSQEQLEHRERSLQTLRDIQRML 25
              :|||:||||||||||||||||:|:|
Db        392 EGLSKEQLEHRERSLQTLRDIERLL 416


RESULT 8
Q67FY2
ID   Q67FY2      PRELIMINARY;      PRT;  1494 AA.
AC   Q67FY2;
DT   25-OCT-2004 (TrEMBLrel. 28, Created)
DT   25-OCT-2004 (TrEMBLrel. 28, Last sequence update)
DT   25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DE   BCL9-2.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6;
RA   Brembeck F.H., Schwarz-Romond T., Bakkers J., Wilhelm S.,
RA   Hammerschmidt M., Birchmeier W.;
RT   "Essential role of BCL9-2 in the switch between [beta]-catenin's
RT   adhesive and transcriptional functions.";
RL   Genes Dev. 18:0-0(2004).
DR   EMBL; AY296058; AAQ62696.1; -.
SQ   SEQUENCE   1494 AA;  156679 MW;  31A9904C5923581C CRC64;
```

```
  Query Match             59.6%;  Score 109;  DB 2;  Length 1494;
  Best Local Similarity   84.0%;  Pred. No. 7.5e-05;
  Matches   21;  Conservative    4;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 DGLSQEQLEHRERSLQTLRDIQRML 25
              :|||:||||||||||||||||:|:|
Db        394 EGLSKEQLEHRERSLQTLRDIERLL 418


RESULT 9
Q6I7B5
ID   Q6I7B5      PRELIMINARY;      PRT;  1494 AA.
AC   Q6I7B5;
DT   05-JUL-2004 (TrEMBLrel. 27, Created)
DT   05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT   05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE   BCL9-related beta-catenin-binding protein.
GN   Name=B9L;
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Adachi S., Jigami T., Yasui T., Nakano T., Ohwada S., Omori Y.,
RA   Sugano S., Ohkawara B., Shibuya H., Nakamura T., Akiyama T.;
RL   Submitted (DEC-2003) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB128033; BAD24964.1; -.
SQ   SEQUENCE   1494 AA;  156570 MW;  71AE96FD33743A6C CRC64;

  Query Match             59.6%;  Score 109;  DB 2;  Length 1494;
  Best Local Similarity   84.0%;  Pred. No. 7.5e-05;
  Matches   21;  Conservative    4;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 DGLSQEQLEHRERSLQTLRDIQRML 25
              :|||:||||||||||||||||:|:|
Db        394 EGLSKEQLEHRERSLQTLRDIERLL 418


RESULT 10
Q86UU0
ID   Q86UU0      PRELIMINARY;      PRT;  1499 AA.
AC   Q86UU0;
DT   01-JUN-2003 (TrEMBLrel. 24, Created)
DT   01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   DLNB11 protein.
GN   Name=DLNB11;
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Kubo T., Arai Y., Ohira M., Gamou T., Maeno G., Sakiyama T.,
RA   Toyoda A., Hattori M., Sakaki Y., Nakagawara A., Ohki M.;
RL   Submitted (OCT-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB094091; BAC76045.1; -.
SQ   SEQUENCE   1499 AA;  157129 MW;  8415C2EDB7AA9C0C CRC64;

  Query Match             59.6%;  Score 109;  DB 2;  Length 1499;
  Best Local Similarity   84.0%;  Pred. No. 7.6e-05;
  Matches   21;  Conservative    4;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 DGLSQEQLEHRERSLQTLRDIQRML 25
              :|||:||||||||||||||||:|:|
Db        397 EGLSKEQLEHRERSLQTLRDIERLL 421


RESULT 11
BCL9_DROME
ID   BCL9_DROME     STANDARD;      PRT;  1469 AA.
AC   Q961D9; Q9V4D2;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   25-JAN-2005 (Rel. 46, Last annotation update)
DE   Bcl-9 homolog (Legless protein).
GN   Name=lgs; Synonyms=BCL9; ORFNames=CG2041;
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC   Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Berkeley;
RX   MEDLINE=20196006; PubMed=10731132; DOI=10.1126/science.287.5461.2185;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D.A., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J.H., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
RN   [2]
RP   GENOME REANNOTATION.
RX   MEDLINE=22426069; PubMed=12537572;
RA   Misra S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,
RA   Hradecky P., Huang Y., Kaminker J.S., Millburn G.H., Prochnik S.E.,
RA   Smith C.D., Tupy J.L., Whitfield E.J., Bayraktaroglu L., Berman B.P.,
RA   Bettencourt B.R., Celniker S.E., de Grey A.D.N.J., Drysdale R.A.,
RA   Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.Q.,
RA   Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,
RA   Lewis S.E.;
RT   "Annotation of the Drosophila melanogaster euchromatic genome: a
RT   systematic review.";
RL   Genome Biol. 3:RESEARCH0083.1-RESEARCH0083.22(2002).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Berkeley; TISSUE=Embryo;
RX   MEDLINE=22426066; PubMed=12537569;
RA   Stapleton M., Carlson J.W., Brokstein P., Yu C., Champe M.,
RA   George R.A., Guarin H., Kronmiller B., Pacleb J.M., Park S., Wan K.H.,
RA   Rubin G.M., Celniker S.E.;
RT   "A Drosophila full-length cDNA resource.";
```

```
RL   Genome Biol. 3:RESEARCH0080.1-RESEARCH0080.8(2002).
RN   [4]
RP   SEQUENCE OF 6-1469 FROM N.A., AND MUTAGENESIS OF GLY-514; LEU-534 AND
RP   ILE-537.
RX   MEDLINE=21952490; PubMed=11955446; DOI=10.1016/S0092-8674(02)00679-7;
RA   Kramps T., Peter O., Brunner E., Nellen D., Froesch B., Chatterjee S.,
RA   Murone M., Zuellig S., Basler K.;
RT   "Wnt/wingless signaling requires BCL9/legless-mediated recruitment of
RT   pygopus to the nuclear beta-catenin-TCF complex.";
RL   Cell 109:47-60(2002).
CC   -!- FUNCTION: Involved in signal transduction through the Wnt pathway.
CC   -!- SUBUNIT: Binds to ARM and PYGO.
CC   -!- SUBCELLULAR LOCATION: Nuclear.
CC   -!- DEVELOPMENTAL STAGE: Expressed both maternally and zygotically
CC       throughout development.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AE003844; AAF59345.2; -.
DR   EMBL; AY051651; AAK93075.1; -.
DR   EMBL; AF457205; AAL91368.1; -.
DR   FlyBase; FBgn0039907; lgs.
DR   GO; GO:0005634; C:nucleus; IDA.
DR   GO; GO:0030528; F:transcription regulator activity; IPI.
DR   GO; GO:0030177; P:positive regulation of Wnt receptor signali. . .; IPI.
DR   GO; GO:0007367; P:segment polarity determination; IMP.
KW   Developmental protein; Nuclear protein; Segmentation polarity protein;
KW   Wnt signaling pathway.
FT   DOMAIN      511    555       ARM-binding.
FT   DOMAIN     1134   1173       Asn-rich.
FT   DOMAIN     1340   1449       Gln-rich.
FT   DOMAIN     1162   1169       Poly-Asn.
FT   MUTAGEN     514    514       G->E: In allele lgs-21L.
FT   MUTAGEN     534    534       L->F: In allele lgs-17E; segment polarity
FT                                phenotype.
FT   MUTAGEN     537    537       I->K: In allele lgs-17P.
SQ   SEQUENCE   1469 AA;  153759 MW;  5672E01B7200ED08 CRC64;

  Query Match             36.1%;  Score 66;  DB 1;  Length 1469;
  Best Local Similarity   31.4%;  Pred. No. 22;
  Matches   11;  Conservative   10;  Mismatches   14;  Indels    0;  Gaps    0;

Qy          1 DGLSQEQLEHRERSLQTLRDIQRMLFPDEKEFTGA 35
              : |: :| :|||  |  :: : : |||: :   ||
Db        520 ENLTPQQRQHREEQLAKIKKMNQFLFPENENSVGA 554


RESULT 12
Q7RLG2
ID   Q7RLG2      PRELIMINARY;      PRT;   192 AA.
AC   Q7RLG2;
DT   01-MAR-2004 (TrEMBLrel. 26, Created)
DT   01-MAR-2004 (TrEMBLrel. 26, Last sequence update)
DT   01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE   Hypothetical protein.
GN   Name=PY02582;
OS   Plasmodium yoelii yoelii.
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX   NCBI_TaxID=73239;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=17XNL;
RX   PubMed=12368865; DOI=10.1038/nature01099;
RA   Carlton J.M., Angiuoli S.V., Suh B.B., Kooij T.W., Pertea M.,
RA   Silva J.C., Ermolaeva M.D., Allen J.E., Selengut J.D., Koo H.L.,
RA   Peterson J.D., Pop M., Kosack D.S., Shumway M.F., Bidwell S.L.,
RA   Shallom S.J., van Aken S.E., Riedmuller S.B., Feldblyum T.V.,
RA   Cho J.K., Quackenbush J., Sedegah M., Shoaibi A., Cummings L.M.,
RA   Florens L., Yates F.R. III, Raine J.D., Sinden R.E., Harris M.A.,
RA   Cunningham D.A., Preiser P.R., Bergman L.W., Vaidya A.B.,
RA   van Lin L.H., Janse C.J., Waters A.P., Smith H.O., White O.R.,
RA   Salzberg S.L., Venter J.C., Fraser C.M., Hoffman S.L., Gardner M.J.,
RA   Carucci D.J.;
RT   "Genome sequence and comparative analysis of the model rodent malaria
RT   parasite Plasmodium yoelii yoelii.";
RL   Nature 419:512-519(2002).
CC   -!- CAUTION: The sequence shown here is derived from an
CC       EMBL/GenBank/DDBJ whole genome shotgun (WGS) entry which is
CC       preliminary data.
DR   EMBL; AABL01000710; EAA22049.1; -.
KW   Hypothetical protein.
SQ   SEQUENCE   192 AA;  22971 MW;  40C958BFC68C6754 CRC64;

  Query Match             33.9%;  Score 62;  DB 2;  Length 192;
  Best Local Similarity   35.5%;  Pred. No. 7.4;
  Matches   11;  Conservative   10;  Mismatches   10;  Indels    0;  Gaps    0;

Qy          1 DGLSQEQLEHRERSLQTLRDIQRMLFPDEKE 31
              : |||::|| :|:    : |: :::|  |||
Db         41 EALSQKKLEEKEKVSDDIYDLYKLVFASEKE 71


RESULT 13
Q89YQ1
ID   Q89YQ1      PRELIMINARY;      PRT;   833 AA.
AC   Q89YQ1;
DT   01-JUN-2003 (TrEMBLrel. 24, Created)
DT   01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT   01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE   DNA mismatch repair protein MutS.
GN   OrderedLocusNames=BT4680;
OS   Bacteroides thetaiotaomicron.
OC   Bacteria; Bacteroidetes; Bacteroides (class); Bacteroidales;
OC   Bacteroidaceae; Bacteroides.
OX   NCBI_TaxID=818;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=VPI-5482 / ATCC 29148;
RX   MEDLINE=22550858; PubMed=12663928; DOI=10.1126/science.1080029;
RA   Xu J., Bjursell M.K., Himrod J., Deng S., Carmichael L.K.,
RA   Chiang H.C., Hooper L.V., Gordon J.I.;
RT   "A genomic view of the human-Bacteroides thetaiotaomicron symbiosis.";
RL   Science 299:2074-2076(2003).
CC   -!- SIMILARITY: Belongs to the DNA mismatch repair mutS family.
DR   EMBL; AE016946; AAO79785.1; -.
DR   GO; GO:0005524; F:ATP binding; IEA.
DR   GO; GO:0003684; F:damaged DNA binding; IEA.
DR   GO; GO:0006259; P:DNA metabolism; IEA.
DR   GO; GO:0006298; P:mismatch repair; IEA.
DR   InterPro; IPR000432; MutS_C.
DR   InterPro; IPR007696; MutS_III.
DR   InterPro; IPR002625; Smr/MutS2_C.
DR   Pfam; PF00488; MutS_V; 1.
DR   Pfam; PF01713; Smr; 1.
DR   ProDom; PD001263; MutS_C; 1.
DR   SMART; SM00534; MUTSac; 1.
DR   SMART; SM00533; MUTSd; 1.
KW   Complete proteome; DNA-binding.
SQ   SEQUENCE   833 AA;  94722 MW;  3B40B0168D6E7076 CRC64;

  Query Match             33.3%;  Score 61;  DB 2;  Length 833;
  Best Local Similarity   44.8%;  Pred. No. 51;
  Matches   13;  Conservative    7;  Mismatches    9;  Indels    0;  Gaps    0;

Qy          3 LSQEQLEHRERSLQTLRDIQRMLFPDEKE 31
              | :::|    |||:|:||| | |  :|:|
Db         87 LDEQELFDLRRSLETIRDIVRFLHRNEEE 115


RESULT 14
```

```
Q7VNS5
ID   Q7VNS5      PRELIMINARY;      PRT;   102 AA.
AC   Q7VNS5;
DT   01-OCT-2003 (TrEMBLrel. 25, Created)
DT   01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT   01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE   Hypothetical protein.
GN   OrderedLocusNames=HD0412;
OS   Haemophilus ducreyi.
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
OC   Pasteurellaceae; Haemophilus.
OX   NCBI_TaxID=730;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=35000HP / ATCC 700724;
RA   Munson R.S. Jr., Ray W.C., Mahairas G., Sabo P., Mungur R.,
RA   Johnson L., Nguyen D., Wang J., Forst C., Hood L.;
RT   "The complete genome sequence of Haemophilus ducreyi.";
RL   Submitted (JUN-2003) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AE017152; AAP95377.1; -.
DR   InterPro; IPR008314; UCP029143.
DR   ProDom; PD030073; UCP029143; 1.
KW   Complete proteome; Hypothetical protein.
SQ   SEQUENCE   102 AA;  12071 MW;  6E2DECFBC5B14256 CRC64;

  Query Match             31.7%;  Score 58;  DB 2;  Length 102;
  Best Local Similarity   31.2%;  Pred. No. 12;
  Matches   10;  Conservative   10;  Mismatches   12;  Indels    0;  Gaps    0;

Qy          1 DGLSQEQLEHRERSLQTLRDIQRMLFPDEKEF 32
              | |:::  :    ||  || : | ::|:|::|
Db         62 DDLNEQHAQQENMSLAELRQVIREIYPNEQKF 93


RESULT 15
Q8Z8Y6
ID   Q8Z8Y6      PRELIMINARY;      PRT;  1034 AA.
AC   Q8Z8Y6;
DT   01-MAR-2002 (TrEMBLrel. 20, Created)
DT   01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Exonuclease SbcC.
GN   Name=sbcC; OrderedLocusNames=STY0429;
OS   Salmonella typhi.
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC   Enterobacteriaceae; Salmonella.
OX   NCBI_TaxID=601;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CT18;
RX   MEDLINE=21534947; PubMed=11677608; DOI=10.1038/35101607;
RA   Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J.,
RA   Churcher C.M., Mungall K.L., Bentley S.D., Holden M.T.G., Sebaihia M.,
RA   Baker S., Basham D., Brooks K., Chillingworth T., Connerton P.,
RA   Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,
RA   Feltwell T., Hamlin N., Haque A., Hien T.T., Holroyd S., Jagels K.,
RA   Krogh A., Larsen T.S., Leather S., Moule S., O'Gaora P., Parry C.,
RA   Quail M.A., Rutherford K.M., Simmonds M., Skelton J., Stevens K.,
RA   Whitehead S., Barrell B.G.;
RT   "Complete genome sequence of a multiple drug resistant Salmonella
RT   enterica serovar Typhi CT18.";
RL   Nature 413:848-852(2001).
DR   EMBL; AL627266; CAD08850.1; -.
DR   GO; GO:0016020; C:membrane; IEA.
DR   GO; GO:0005634; C:nucleus; IEA.
DR   GO; GO:0005524; F:ATP binding; IEA.
DR   GO; GO:0042626; F:ATPase activity, coupled to transmembrane m. . .; IEA.
DR   GO; GO:0004527; F:exonuclease activity; IEA.
DR   GO; GO:0007059; P:chromosome segregation; IEA.
DR   GO; GO:0006259; P:DNA metabolism; IEA.
DR   GO; GO:0006810; P:transport; IEA.
DR   Pfam; PF02463; SMC_N; 1.
DR   TIGRFAMs; TIGR00618; sbcc; 1.
KW   Complete proteome; Exonuclease.
SQ   SEQUENCE   1034 AA;  116759 MW;  2513B7573626960A CRC64;

  Query Match             31.1%;  Score 57;  DB 2;  Length 1034;
  Best Local Similarity   42.9%;  Pred. No. 2.1e+02;
  Matches   12;  Conservative    6;  Mismatches   10;  Indels    0;  Gaps    0;

Qy          3 LSQEQLEHRERSLQTLRDIQRMLFPDEK 30
              |: |||:  | ||| | | :: |  |::
Db        213 LADEQLQQLEASLQALTDEEKRLLADQQ 240


RESULT 16
Q83SZ9
ID   Q83SZ9      PRELIMINARY;      PRT;  1036 AA.
AC   Q83SZ9;
DT   01-JUN-2003 (TrEMBLrel. 24, Created)
DT   01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Exonuclease SbcC.
GN   Name=sbcC; OrderedLocusNames=t2469;
OS   Salmonella typhi.
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC   Enterobacteriaceae; Salmonella.
OX   NCBI_TaxID=601;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Ty2 / ATCC 700931;
RX   MEDLINE=22531367; PubMed=12644504;
RX   DOI=10.1128/JB.185.7.2330-2337.2003;
RA   Deng W., Liou S.-R., Plunkett G. III, Mayhew G.F., Rose D.J.,
RA   Burland V., Kodoyianni V., Schwartz D.C., Blattner F.R.;
RT   "Comparative genomics of Salmonella enterica serovar Typhi strains Ty2
RT   and CT18.";
RL   J. Bacteriol. 185:2330-2337(2003).
DR   EMBL; AE016842; AAO70059.1; -.
DR   GO; GO:0016020; C:membrane; IEA.
DR   GO; GO:0005634; C:nucleus; IEA.
DR   GO; GO:0005524; F:ATP binding; IEA.
DR   GO; GO:0042626; F:ATPase activity, coupled to transmembrane m. . .; IEA.
DR   GO; GO:0004527; F:exonuclease activity; IEA.
DR   GO; GO:0007059; P:chromosome segregation; IEA.
DR   GO; GO:0006259; P:DNA metabolism; IEA.
DR   GO; GO:0006810; P:transport; IEA.
DR   InterPro; IPR003439; ABC_transporter.
DR   InterPro; IPR004592; SbcC.
DR   InterPro; IPR003395; SMC_N.
DR   Pfam; PF02463; SMC_N; 1.
DR   TIGRFAMs; TIGR00618; sbcc; 1.
KW   Exonuclease.
SQ   SEQUENCE   1036 AA;  116943 MW;  31EE32204ADBA905 CRC64;

  Query Match             31.1%;  Score 57;  DB 2;  Length 1036;
  Best Local Similarity   42.9%;  Pred. No. 2.1e+02;
  Matches   12;  Conservative    6;  Mismatches   10;  Indels    0;  Gaps    0;

Qy          3 LSQEQLEHRERSLQTLRDIQRMLFPDEK 30
              |: |||:  | ||| | | :: |  |::
Db        213 LADEQLQQLEASLQALTDEEKRLLADQQ 240


RESULT 17
Q8ZRE3
ID   Q8ZRE3      PRELIMINARY;      PRT;  1046 AA.
AC   Q8ZRE3;
DT   01-MAR-2002 (TrEMBLrel. 20, Created)
DT   01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   ATP-dependent dsDNA exonuclease.
GN   Name=sbcC; OrderedLocusNames=STM0395;
OS   Salmonella typhimurium.
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC   Enterobacteriaceae; Salmonella.
```

```
OX   NCBI_TaxID=602;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=LT2;
RX   MEDLINE=21534948; PubMed=11677609; DOI=10.1038/35101614;
RA   McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
RA   Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,
RA   Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
RA   Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
RA   Waterston R., Wilson R.K.;
RT   "Complete genome sequence of Salmonella enterica serovar Typhimurium
RT   LT2.";
RL   Nature 413:852-856(2001).
DR   EMBL; AE008713; AAL19349.1; -.
DR   GO; GO:0016020; C:membrane; IEA.
DR   GO; GO:0005634; C:nucleus; IEA.
DR   GO; GO:0005524; F:ATP binding; IEA.
DR   GO; GO:0042626; F:ATPase activity, coupled to transmembrane m. . .; IEA.
DR   GO; GO:0004527; F:exonuclease activity; IEA.
DR   GO; GO:0007059; P:chromosome segregation; IEA.
DR   GO; GO:0006259; P:DNA metabolism; IEA.
DR   GO; GO:0006810; P:transport; IEA.
DR   Pfam; PF02463; SMC_N; 1.
DR   TIGRFAMs; TIGR00618; sbcc; 1.
KW   Complete proteome; Exonuclease.
SQ   SEQUENCE   1046 AA;  117823 MW;  BA565CA3BDAD0C82 CRC64;

  Query Match             31.1%;  Score 57;  DB 2;  Length 1046;
  Best Local Similarity   42.9%;  Pred. No. 2.1e+02;
  Matches   12;  Conservative    6;  Mismatches   10;  Indels    0;  Gaps    0;

Qy          3 LSQEQLEHRERSLQTLRDIQRMLFPDEK 30
              |: |||:  | ||| | | :: |  |::
Db        213 LADEQLQQLEASLQALTDEEKRLLADQQ 240


RESULT 18
Q7Q3P6
ID   Q7Q3P6      PRELIMINARY;      PRT;   476 AA.
AC   Q7Q3P6;
DT   01-MAR-2004 (TrEMBLrel. 26, Created)
DT   01-MAR-2004 (TrEMBLrel. 26, Last sequence update)
DT   01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE   AgCP10924 (Fragment).
GN   Name=agCG50252; ORFNames=ENSANGG00000009382;
OS   Anopheles gambiae str. PEST.
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC   Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea; Anopheles.
OX   NCBI_TaxID=180454;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=PEST;
RA   Anopheles Genome Sequencing Consortium;
RL   Submitted (MAR-2002) to the EMBL/GenBank/DDBJ databases.
CC   -!- CAUTION: The sequence shown here is derived from an
CC       EMBL/GenBank/DDBJ whole genome shotgun (WGS) entry which is
CC       preliminary data.
DR   EMBL; AAAB01008964; EAA12408.1; -.
DR   InterPro; IPR009060; UBA_like.
DR   InterPro; IPR001012; UBX.
DR   PROSITE; PS50033; UBX; 1.
FT   NON_TER       1      1
SQ   SEQUENCE   476 AA;  54397 MW;  5A7EC8C1E8C30576 CRC64;

  Query Match             30.9%;  Score 56.5;  DB 2;  Length 476;
  Best Local Similarity   43.8%;  Pred. No. 1e+02;
  Matches   14;  Conservative    3;  Mismatches   10;  Indels    5;  Gaps    1;

Qy          7 QLEHRERSLQTLRDIQRMLF-----PDEKEFT 33
              :|| | ||  |:|||   :|     ||  | |
Db        408 RLERRFRSTDTMRDIYHFIFCHPDAPDSFEIT 439


RESULT 19
TRF4_YEAST
ID   TRF4_YEAST     STANDARD;      PRT;   584 AA.
AC   P53632;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   25-OCT-2004 (Rel. 45, Last annotation update)
DE   Topoisomerase 1-related protein TRF4.
GN   Name=TRF4; OrderedLocusNames=YOL115W; ORFNames=O0716, HRC584;
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=96109595; PubMed=8647385;
RA   Sadoff B.U., Heath-Pagliuso S., Castano I.B., Zhu Y., Kieff F.S.,
RA   Christman M.F.;
RT   "Isolation of mutants of Saccharomyces cerevisiae requiring DNA
RT   topoisomerase I.";
RL   Genetics 141:465-479(1995).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=96076631; PubMed=7502582;
RA   Vandenbol M., Durand P., Portetelle D., Hilger F.;
RT   "Sequence analysis of a 44 kb DNA fragment of yeast chromosome XV
RT   including the Tyl-H3 retrotransposon, the suf1(+) frameshift
RT   suppressor gene for tRNA-Gly, the yeast transfer RNA-Thr-1a and a
RT   delta element.";
RL   Yeast 11:1069-1075(1995).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288c;
RA   Marsischky G., Rolfs A., Richardson A., Kane M., Baqui M., Taycher E.,
RA   Hu Y., Vannberg F., Weger J., Kramer J., Moreira D., Kelley F.,
RA   Zuo D., Raphael J., Hogle C., Jepson D., Williamson J., Camargo A.,
RA   Gonzaga L., Vasconcelos A.T., Simpson A., Kolodner R., Harlow E.,
RA   LaBaer J.;
RT   "Creation of the YFLEX clone resource: cloning of Saccharomyces
RT   cerevisiae ORFs in the Gateway recombinational cloning system.";
RL   Submitted (AUG-2004) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: Essential protein required for proper nuclear division
CC       in mitosis. May mediate mitotic chromosome condensation.
CC   -!- SIMILARITY: Belongs to the CID1/TRF4/TRF5 family.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U31355; AAC49091.1; -.
DR   EMBL; Z48149; CAA88145.1; -.
DR   EMBL; Z74857; CAA99134.1; -.
DR   EMBL; AY723865; AAU09782.1; -.
DR   PIR; S51882; S51882.
DR   GermOnline; 143537; -.
DR   SGD; S000005475; TRF4.
DR   GO; GO:0005634; C:nucleus; IDA.
DR   GO; GO:0003887; F:DNA-directed DNA polymerase activity; IDA.
DR   GO; GO:0006265; P:DNA topological change; IGI.
DR   GO; GO:0007076; P:mitotic chromosome condensation; IMP.
DR   InterPro; IPR002934; NTP_transf.
DR   InterPro; IPR001201; PAP_25A_core.
DR   InterPro; IPR002058; PAP_assoc.
DR   Pfam; PF01909; NTP_transf_2; 1.
DR   Pfam; PF03828; PAP_assoc; 1.
KW   Mitosis.
SQ   SEQUENCE   584 AA;  66030 MW;  8A58B29E4BFDC022 CRC64;

  Query Match             30.6%;  Score 56;  DB 1;  Length 584;
```

```
  Best Local Similarity   36.0%;  Pred. No. 1.5e+02;
  Matches    9;  Conservative   10;  Mismatches    6;  Indels    0;  Gaps    0;

Qy          4 SQEQLEHRERSLQTLRDIQRMLFPD 28
              |:|::| | ::: |:|:  : |:||
Db        193 SREEIEIRNQTISTIREAVKQLWPD 217


RESULT 20
Q8TGZ1
ID   Q8TGZ1      PRELIMINARY;      PRT;   818 AA.
AC   Q8TGZ1;
DT   01-JUN-2002 (TrEMBLrel. 21, Created)
DT   01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Archaea-specific Superfamily II helicase.
GN   OrderedLocusNames=MK0835;
OS   Methanopyrus kandleri.
OC   Archaea; Euryarchaeota; Methanopyri; Methanopyrales; Methanopyraceae;
OC   Methanopyrus.
OX   NCBI_TaxID=2320;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=AV19 / DSM 6324 / JCM 9639;
RX   MEDLINE=21927647; PubMed=11930014; DOI=10.1073/pnas.032671499;
RA   Slesarev A.I., Mezhevaya K.V., Makarova K.S., Polushin N.N.,
RA   Shcherbinina O.V., Shakhova V.V., Belova G.I., Aravind L.,
RA   Natale D.A., Rogozin I.B., Tatusov R.L., Wolf Y.I., Stetter K.O.,
RA   Malykh A.G., Koonin E.V., Kozyavkin S.A.;
RT   "The complete genome of hyperthermophile Methanopyrus kandleri AV19
RT   and monophyly of archaeal methanogens.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:4644-4649(2002).
DR   EMBL; AE010374; AAM02048.1; -.
DR   GO; GO:0005524; F:ATP binding; IEA.
DR   GO; GO:0008026; F:ATP-dependent helicase activity; IEA.
DR   GO; GO:0003676; F:nucleic acid binding; IEA.
DR   GO; GO:0000166; F:nucleotide binding; IEA.
DR   InterPro; IPR003593; AAA_ATPase.
DR   InterPro; IPR001410; DEAD.
DR   InterPro; IPR011545; DEAD/DEAH_N.
DR   InterPro; IPR001650; Helicase_C.
DR   Pfam; PF00270; DEAD; 1.
DR   Pfam; PF00271; Helicase_C; 1.
DR   SMART; SM00382; AAA; 1.
DR   SMART; SM00487; DEXDc; 1.
DR   SMART; SM00490; HELICc; 1.
KW   ATP-binding; Complete proteome; Helicase; Hydrolase.
SQ   SEQUENCE   818 AA;  91715 MW;  C2136200A710817E CRC64;

  Query Match             30.6%;  Score 56;  DB 2;  Length 818;
  Best Local Similarity   42.9%;  Pred. No. 2.1e+02;
  Matches   15;  Conservative    4;  Mismatches   16;  Indels    0;  Gaps    0;

Qy          2 GLSQEQLEHRERSLQTLRDIQRMLFPDEKEFTGAQ 36
              | ||  ||  || |  |||| |::   :  |  |:
Db        134 GFSQTLLEKLERLLHELRDIDRVVEMVDPAFDPAE 168


RESULT 21
Q9NT51
ID   Q9NT51      PRELIMINARY;      PRT;   859 AA.
AC   Q9NT51;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   Hypothetical protein DKFZp434P1818 (Fragment).
GN   Name=DKFZp434P1818;
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Testis;
RG   The German cDNA Consortium;
RA   Ottenwaelder B., Obermaier B., Deutschenbaur S., Schaipp A.,
RA   Mewes H.W., Weil B., Amid C., Osanger A., Fobo G., Han M., Wiemann S.;
RL   Submitted (SEP-2004) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AL137528; CAB70792.1; -.
DR   PIR; T46372; T46372.
DR   InterPro; IPR002013; Syja_N.
DR   Pfam; PF02383; Syja_N; 1.
DR   PROSITE; PS50275; SAC; 1.
KW   Hypothetical protein.
FT   NON_TER       1       1
SQ   SEQUENCE   859 AA;  96781 MW;  BCBBC47C8B726D76 CRC64;

  Query Match             30.6%;  Score 56;  DB 2;  Length 859;
  Best Local Similarity   33.3%;  Pred. No. 2.3e+02;
  Matches   10;  Conservative    8;  Mismatches   12;  Indels    0;  Gaps    0;

Qy          5 QEQLEHRERSLQTLRDIQRMLFPDEKEFTG 34
              : |  |:|   | |:   ::| ||:::| |
Db        317 ENQRSHQELISQLLQSYMKLLLPDDEKFHG 346


RESULT 22
Q9Y2H2
ID   Q9Y2H2      PRELIMINARY;      PRT;  1150 AA.
AC   Q9Y2H2;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-MAR-2004 (TrEMBLrel. 26, Last sequence update)
DT   01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE   KIAA0966 protein (Fragment).
GN   Name=KIAA0966;
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Brain;
RX   MEDLINE=99246063; PubMed=10231032;
RA   Nagase T., Ishikawa K., Suyama M., Kikuno R., Hirosawa M.,
RA   Miyajima N., Tanaka A., Kotani H., Nomura N., Ohara O.;
RT   "Prediction of the coding sequences of unidentified human genes. XIII.
RT   The complete sequences of 100 new cDNA clones from brain which code
RT   for large proteins in vitro.";
RL   DNA Res. 6:63-70(1999).
DR   EMBL; AB023183; BAA76810.2; -.
DR   Genew; HGNC:17054; INPP5F.
DR   InterPro; IPR002013; Syja_N.
DR   Pfam; PF02383; Syja_N; 1.
DR   PROSITE; PS50275; SAC; 1.
FT   NON_TER       1       1
SQ   SEQUENCE   1150 AA;  130165 MW;  A75EB0F636542A13 CRC64;

  Query Match             30.6%;  Score 56;  DB 2;  Length 1150;
  Best Local Similarity   33.3%;  Pred. No. 3.1e+02;
  Matches   10;  Conservative    8;  Mismatches   12;  Indels    0;  Gaps    0;

Qy          5 QEQLEHRERSLQTLRDIQRMLFPDEKEFTG 34
              : |  |:|   | |:   ::| ||:::| |
Db        608 ENQRSHQELISQLLQSYMKLLLPDDEKFHG 637


RESULT 23
Q9VXU1
ID   Q9VXU1      PRELIMINARY;      PRT;  1208 AA.
AC   Q9VXU1;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT   01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE   CG33206-PB.
GN   ORFNames=CG33206;
OS   Drosophila melanogaster (Fruit fly).
```

```
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC   Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20196006; PubMed=10731132; DOI=10.1126/science.287.5461.2185;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.H., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Gabor G.L.,
RA   Abril J.F., Agbayani A., An H.J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., WoodageT, Worley K.C., Wu D., Yang S., Yao Q.A., Ye J.,
RA   Yeh R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=22426065; PubMed=12537568;
RA   Celniker S.E., Wheeler D.A., Kronmiller B., Carlson J.W., Halpern A.,
RA   Patel S., Adams M., Champe M., Dugan S.P., Frise E., Hodgson A.,
RA   George R.A., Hoskins R.A., Laverty T., Muzny D.M., Nelson C.R.,
RA   Pacleb J.M., Park S., Pfeiffer B.D., Richards S., Sodergren E.J.,
RA   Svirskas R., Tabor P.E., Wan K., Stapleton M., Sutton G.G., Venter C.,
RA   Weinstock G., Scherer S.E., Myers E.W., Gibbs R.A., Rubin G.M.;
RT   "Finishing a whole-genome shotgun: Release 3 of the Drosophila
RT   melanogaster euchromatic genome sequence.";
RL   Genome Biol. 3:RESEARCH0079-RESEARCH0079(2002).
RN   [3]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=22426070; PubMed=12537573;
RA   Kaminker J.S., Bergman C.M., Kronmiller B., Carlson J., Svirskas R.,
RA   Patel S., Frise E., Wheeler D.A., Lewis S.E., Rubin G.M.,
RA   Ashburner M., Celniker S.E.;
RT   "The transposable elements of the Drosophila melanogaster euchromatin:
RT   a genomics perspective.";
RL   Genome Biol. 3:RESEARCH0084-RESEARCH0084(2002).
RN   [4]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=22426069; PubMed=12537572;
RA   Misra S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,
RA   Hradecky P., Huang Y., Kaminker J.S., Millburn G.H., Prochnik S.E.,
RA   Smith C.D., Tupy J.L., Whitfied E.J., Bayraktaroglu L., Berman B.P.,
RA   Bettencourt B.R., Celniker S.E., de Grey A.D., Drysdale R.A.,
RA   Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.Q.,
RA   Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,
RA   Lewis S.E.;
RT   "Annotation of the Drosophila melanogaster euchromatic genome: a
RT   systematic review.";
RL   Genome Biol. 3:RESEARCH0083-RESEARCH0083(2002).
RN   [5]
RP   SEQUENCE FROM N.A.
RG   FlyBase;
RL   Submitted (SEP-2002) to the EMBL/GenBank/DDBJ databases.
RN   [6]
RP   SEQUENCE FROM N.A.
RG   FlyBase;
RL   Submitted (MAR-2004) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AE003499; AAF48467.2; -.
DR   FlyBase; FBgn0027287; l(1)G0168.
DR   InterPro; IPR000237; GRIP.
DR   PROSITE; PS50913; GRIP; 1.
SQ   SEQUENCE   1208 AA;  137773 MW;  131F2EB663826A92 CRC64;

  Query Match             30.6%;  Score 56;  DB 2;  Length 1208;
  Best Local Similarity   45.5%;  Pred. No. 3.3e+02;
  Matches   10;  Conservative    6;  Mismatches    6;  Indels    0;  Gaps    0;

Qy          3 LSQEQLEHRERSLQTLRDIQRM 24
              | |:| | :|:   ||||::|:
Db        811 LQQQQAESQEQQASTLRDLERL 832


RESULT 24
Q9VXU2
ID   Q9VXU2      PRELIMINARY;      PRT;  1398 AA.
AC   Q9VXU2; Q960D0;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT   25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DE   CG33206-PA (SD07366p).
GN   Name=CG6350; ORFNames=CG33206;
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC   Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20196006; PubMed=10731132; DOI=10.1126/science.287.5461.2185;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.H., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Gabor G.L.,
RA   Abril J.F., Agbayani A., An H.J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
```

```
RA   Reinert K., Remington K., Saunders R.D., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., WoodageT, Worley K.C., Wu D., Yang S., Yao Q.A., Ye J.,
RA   Yeh R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=22426065; PubMed=12537568;
RA   Celniker S.E., Wheeler D.A., Kronmiller B., Carlson J.W., Halpern A.,
RA   Patel S., Adams M., Champe M., Dugan S.P., Frise E., Hodgson A.,
RA   George R.A., Hoskins R.A., Laverty T., Muzny D.M., Nelson C.R.,
RA   Pacleb J.M., Park S., Pfeiffer B.D., Richards S., Sodergren E.J.,
RA   Svirskas R., Tabor P.E., Wan K., Stapleton M., Sutton G.G., Venter C.,
RA   Weinstock G., Scherer S.E., Myers E.W., Gibbs R.A., Rubin G.M.;
RT   "Finishing a whole-genome shotgun: Release 3 of the Drosophila
RT   melanogaster euchromatic genome sequence.";
RL   Genome Biol. 3:RESEARCH0079-RESEARCH0079(2002).
RN   [3]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=22426070; PubMed=12537573;
RA   Kaminker J.S., Bergman C.M., Kronmiller B., Carlson J., Svirskas R.,
RA   Patel S., Frise E., Wheeler D.A., Lewis S.E., Rubin G.M.,
RA   Ashburner M., Celniker S.E.;
RT   "The transposable elements of the Drosophila melanogaster euchromatin:
RT   a genomics perspective.";
RL   Genome Biol. 3:RESEARCH0084-RESEARCH0084(2002).
RN   [4]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=22426069; PubMed=12537572;
RA   Misra S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,
RA   Hradecky P., Huang Y., Kaminker J.S., Millburn G.H., Prochnik S.E.,
RA   Smith C.D., Tupy J.L., Whitfied E.J., Bayraktaroglu L., Berman B.P.,
RA   Bettencourt B.R., Celniker S.E., de Grey A.D., Drysdale R.A.,
RA   Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.Q.,
RA   Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,
RA   Lewis S.E.;
RT   "Annotation of the Drosophila melanogaster euchromatic genome: a
RT   systematic review.";
RL   Genome Biol. 3:RESEARCH0083-RESEARCH0083(2002).
RN   [5]
RP   SEQUENCE FROM N.A.
RG   FlyBase;
RL   Submitted (SEP-2002) to the EMBL/GenBank/DDBJ databases.
RN   [6]
RP   SEQUENCE FROM N.A.
RG   FlyBase;
RL   Submitted (MAR-2004) to the EMBL/GenBank/DDBJ databases.
RN   [7]
RP   SEQUENCE FROM N.A.
RA   Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
RA   Champe M., Chavez C., Dorsett V., Dresnek D., Farfan D., Frise E.,
RA   George R., Gonzalez M., Guarin H., Kronmiller B., Li P., Liao G.,
RA   Miranda A., Mungall C.J., Nunoo J., Pacleb J., Paragas V., Park S.,
RA   Patel S., Phouanenavong S., Wan K., Yu C., Lewis S.E., Rubin G.M.,
RA   Celniker S.;
RL   Submitted (JAN-2003) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AE003499; AAF48466.2; -.
DR   EMBL; AY052121; AAK93545.1; -.
DR   FlyBase; FBgn0027287; l(1)G0168.
DR   InterPro; IPR000237; GRIP.
DR   PROSITE; PS50913; GRIP; 1.
SQ   SEQUENCE   1398 AA;  158483 MW;  8BA1C2FE3E9F555D CRC64;

  Query Match             30.6%;  Score 56;  DB 2;  Length 1398;
  Best Local Similarity   45.5%;  Pred. No. 3.9e+02;
  Matches   10;  Conservative    6;  Mismatches    6;  Indels    0;  Gaps    0;

Qy          3 LSQEQLEHRERSLQTLRDIQRM 24
              | |:| | :|:   ||||::|:
Db       1001 LQQQQAESQEQQASTLRDLERL 1022


RESULT 25
Q81U73
ID   Q81U73      PRELIMINARY;      PRT;   767 AA.
AC   Q81U73; Q6I2G0; Q6KW87;
DT   01-JUN-2003 (TrEMBLrel. 24, Created)
DT   01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT   25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DE   Hypothetical protein.
GN   OrderedLocusNames=BA1011, BAS0946, GBAA1011;
OS   Bacillus anthracis.
OC   Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX   NCBI_TaxID=1392;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Ames / isolate Porton;
RX   MEDLINE=22608414; PubMed=12721629; DOI=10.1038/nature01586;
RA   Read T.D., Peterson S.N., Tourasse N.J., Baillie L.W., Paulsen I.T.,
RA   Nelson K.E., Tettelin H., Fouts D.E., Eisen J.A., Gill S.R.,
RA   Holtzapple E.K., Okstad O.A., Helgason E., Rilstone J., Wu M.,
RA   Kolonay J.F., Beanan M.J., Dodson R.J., Brinkac L.M., Gwinn M.L.,
RA   DeBoy R.T., Madpu R., Daugherty S.C., Durkin A.S., Haft D.H.,
RA   Nelson W.C., Peterson J.D., Pop M., Khouri H.M., Radune D.,
RA   Benton J.L., Mahamoud Y., Jiang L., Hance I.R., Weidman J.F.,
RA   Berry K.J., Plaut R.D., Wolf A.M., Watkins K.L., Nierman W.C.,
RA   Hazen A., Cline R.T., Redmond C., Thwaite J.E., White O.,
RA   Salzberg S.L., Thomason B., Friedlander A.M., Koehler T.M.,
RA   Hanna P.C., Kolstoe A.-B., Fraser C.M.;
RT   "The genome sequence of Bacillus anthracis Ames and comparison to
RT   closely related bacteria.";
RL   Nature 423:81-86(2003).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Ames / isolate 0581;
RA   Ravel J., Rasko D.A., Shumway M.F., Jiang L., Cer R.Z., Federova N.B.,
RA   Wilson M., Stanley S., Decker S., Read T.D., Salzberg S.L.,
RA   Fraser C.M.;
RT   "Bacillus anthracis comparative genomics.";
RL   Submitted (MAY-2004) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Sterne;
RA   Brettin T.S., Bruce D., Challacombe J.F., Gilna P., Han C., Hill K.,
RA   Hitchcock P., Jackson P., Keim P., Longmire J., Lucas S., Okinaka R.,
RA   Richardson P., Rubin E., Tice H.;
RL   Submitted (JAN-2004) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AE017027; AAP24998.1; -.
DR   EMBL; AE017334; AAT30114.2; -.
DR   EMBL; AE017225; AAT53271.1; -.
DR   TIGR; BA1011; -.
DR   TIGR; GBAA1011; -.
KW   Complete proteome; Hypothetical protein.
SQ   SEQUENCE   767 AA;  90630 MW;  37E501351906E9E4 CRC64;

  Query Match             30.3%;  Score 55.5;  DB 2;  Length 767;
  Best Local Similarity   38.7%;  Pred. No. 2.3e+02;
  Matches   12;  Conservative    8;  Mismatches    8;  Indels    3;  Gaps    1;

Qy          3 LSQEQLEHRERSLQTLRDIQRMLFPDEKEFT 33
              ::||||| :|   : :| ||: :  ||:  |
Db        190 VAQEQLEEQE---ENIRQIQKQMLADEERNT 217


RESULT 26
Q63EZ1
ID   Q63EZ1      PRELIMINARY;      PRT;   974 AA.
AC   Q63EZ1;
DT   25-OCT-2004 (TrEMBLrel. 28, Created)
DT   25-OCT-2004 (TrEMBLrel. 28, Last sequence update)
```

```
DT   25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DE   Hypothetical protein yhaN.
GN   Name=yhaN; ORFNames=BTZK0919;
OS   Bacillus cereus ZK.
OC   Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX   NCBI_TaxID=288681;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ZK;
RA   Brettin T.S., Bruce D., Challacombe J.F., Gilna P., Han C., Hill K.,
RA   Hitchcock P., Jackson P., Keim P., Longmire J., Lucas S., Okinaka R.,
RA   Richardson P., Rubin E., Tice H.;
RT   "Complete genome sequence of Bacillus cereus ZK.";
RL   Submitted (JUL-2004) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; CP000001; AAU19326.1; -.
KW   Hypothetical protein.
SQ   SEQUENCE   974 AA;  114224 MW;  66DA2EE4B43CE41A CRC64;

  Query Match             30.3%;  Score 55.5;  DB 2;  Length 974;
  Best Local Similarity   38.7%;  Pred. No. 3e+02;
  Matches   12;  Conservative    8;  Mismatches    8;  Indels    3;  Gaps    1;

Qy          3 LSQEQLEHRERSLQTLRDIQRMLFPDEKEFT 33
              ::|||||  :|   : :| ||: :  ||:  |
Db        397 VAQEQLEEQE---ENIRQIQKQMLADEERNT 424


RESULT 27
Q8UCA0
ID   Q8UCA0      PRELIMINARY;      PRT;   287 AA.
AC   Q8UCA0;
DT   01-JUN-2002 (TrEMBLrel. 21, Created)
DT   01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT   01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE   Transcriptional regulator, RpiR family.
GN   OrderedLocusNames=Atu2598;
OS   Agrobacterium tumefaciens (strain C58 / ATCC 33970).
OC   Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC   Rhizobiaceae; Rhizobium/Agrobacterium group; Agrobacterium.
OX   NCBI_TaxID=176299;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Dupont;
RX   MEDLINE=21608550; PubMed=11743193; DOI=10.1126/science.1066804;
RA   Wood D.W., Setubal J.C., Kaul R., Monks D.E., Kitajima J.P.,
RA   Okura V.K., Zhou Y., Chen L., Wood G.E., Almeida N.F. Jr., Woo L.,
RA   Chen Y., Paulsen I.T., Eisen J.A., Karp P.D., Bovee D. Sr.,
RA   Chapman P., Clendenning J., Deatherage G., Gillet W., Grant C.,
RA   Kutyavin T., Levy R., Li M.-J., McClelland E., Palmieri A.,
RA   Raymond C., Rouse G., Saenphimmachak C., Wu Z., Romero P., Gordon D.,
RA   Zhang S., Yoo H., Tao Y., Biddle P., Jung M., Krespan W., Perry M.,
RA   Gordon-Kamm B., Liao L., Kim S., Hendrick C., Zhao Z.-Y., Dolan M.,
RA   Chumley F., Tingey S.V., Tomb J.-F., Gordon M.P., Olson M.V.,
RA   Nester E.W.;
RT   "The genome of the natural genetic engineer Agrobacterium tumefaciens
RT   C58.";
RL   Science 294:2317-2323(2001).
DR   EMBL; AE009206; AAL43579.1; -.
DR   PIR; AE2895; AE2895.
DR   PIR; H97670; H97670.
DR   GO; GO:0005529; F:sugar binding; IEA.
DR   GO; GO:0003700; F:transcription factor activity; IEA.
DR   GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR   GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR   Pfam; PF01380; SIS; 1.
DR   PROSITE; PS00356; HTH_LACI_1; UNKNOWN_1.
KW   Complete proteome.
SQ   SEQUENCE   287 AA;  31331 MW;  2877CC6B531A0D4A CRC64;

  Query Match             30.1%;  Score 55;  DB 2;  Length 287;
  Best Local Similarity   40.0%;  Pred. No. 90;
  Matches   12;  Conservative    7;  Mismatches    7;  Indels    4;  Gaps    1;

Qy          9 EHRERSLQTLRDIQRMLF----PDEKEFTG 34
              : |:||: ||| |:: |     ||:|:  |
Db        257 QQRQRSMVTLRHIKQQLVEHRDPDDKQLLG 286


RESULT 28
Q7CWN9
ID   Q7CWN9      PRELIMINARY;      PRT;   295 AA.
AC   Q7CWN9;
DT   05-JUL-2004 (TrEMBLrel. 27, Created)
DT   05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT   05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE   AGR_C_4707p.
GN   OrderedLocusNames=AGR_C_4707;
OS   Agrobacterium tumefaciens (strain C58 / ATCC 33970).
OC   Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC   Rhizobiaceae; Rhizobium/Agrobacterium group; Agrobacterium.
OX   NCBI_TaxID=176299;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Cereon;
RX   MEDLINE=21608551; PubMed=11743194; DOI=10.1126/science.1066803;
RA   Goodner B., Hinkle G., Gattung S., Miller N., Blanchard M.,
RA   Qurollo B., Goldman B.S., Cao Y., Askenazi M., Halling C., Mullin L.,
RA   Houmiel K., Gordon J., Vaudin M., Iartchouk O., Epp A., Liu F.,
RA   Wollam C., Allinger M., Doughty D., Scott C., Lappas C., Markelz B.,
RA   Flanagan C., Crowell C., Gurson J., Lomo C., Sear C., Strub G.,
RA   Cielo C., Slater S.;
RT   "Genome sequence of the plant pathogen and biotechnology agent
RT   Agrobacterium tumefaciens C58.";
RL   Science 294:2323-2328(2001).
DR   EMBL; AE008173; AAK88321.1; -.
DR   GO; GO:0005529; F:sugar binding; IEA.
DR   GO; GO:0003700; F:transcription factor activity; IEA.
DR   GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR   GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR   InterPro; IPR000843; HTH_LacI.
DR   InterPro; IPR001347; SIS.
DR   InterPro; IPR009058; Wing_hlx_DNA_bnd.
DR   Pfam; PF01380; SIS; 1.
DR   PROSITE; PS00356; HTH_LACI_1; UNKNOWN_1.
SQ   SEQUENCE   295 AA;  32258 MW;  3B50926B3CB72456 CRC64;

  Query Match             30.1%;  Score 55;  DB 2;  Length 295;
  Best Local Similarity   40.0%;  Pred. No. 92;
  Matches   12;  Conservative    7;  Mismatches    7;  Indels    4;  Gaps    1;

Qy          9 EHRERSLQTLRDIQRMLF----PDEKEFTG 34
              : |:||: ||| |:: |     ||:|:  |
Db        265 QQRQRSMVTLRHIKQQLVEHRDPDDKQLLG 294


RESULT 29
Q6PCC7
ID   Q6PCC7      PRELIMINARY;      PRT;   335 AA.
AC   Q6PCC7;
DT   05-JUL-2004 (TrEMBLrel. 27, Created)
DT   05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT   05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE   Hypothetical protein (Fragment).
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Placenta;
RX   MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
```

```
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA   Krzywinski M.I., Skalska U., Smailus D.E., Schnerch A., Schein J.E.,
RA   Jones S.J., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length human
RT   and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Placenta;
RA   Strausberg R.;
RL   Submitted (OCT-2003) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC059382; AAH59382.1; -.
DR   HSSP; Q07960; 1AM4.
DR   InterPro; IPR001251; CRAL_TRIO_C.
DR   InterPro; IPR000198; RhoGAP.
DR   InterPro; IPR008936; Rho_GAP.
DR   Pfam; PF00620; RhoGAP; 1.
DR   SMART; SM00324; RhoGAP; 1.
DR   SMART; SM00516; SEC14; 1.
DR   PROSITE; PS50191; CRAL_TRIO; 1.
DR   PROSITE; PS50238; RHOGAP; 1.
KW   Hypothetical protein.
FT   NON_TER       1      1
SQ   SEQUENCE   335 AA;  38646 MW;  7B1B179BD5873F9A CRC64;

  Query Match             30.1%;  Score 55;  DB 2;  Length 335;
  Best Local Similarity   52.2%;  Pred. No. 1.1e+02;
  Matches   12;  Conservative    4;  Mismatches    7;  Indels    0;  Gaps    0;

Qy          2 GLSQEQLEHRERSLQTLRDIQRM 24
              ||  | |  |  |:||:|:|||:
Db        265 GLRTEGLFRRSASVQTVREIQRL 287


RESULT 30
Q6NVX9
ID   Q6NVX9      PRELIMINARY;      PRT;   337 AA.
AC   Q6NVX9;
DT   05-JUL-2004 (TrEMBLrel. 27, Created)
DT   05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT   05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE   ARHGAP8 protein (Fragment).
GN   Name=ARHGAP8;
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Placenta;
RX   MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA   Krzywinski M.I., Skalska U., Smailus D.E., Schnerch A., Schein J.E.,
RA   Jones S.J., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length human
RT   and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Placenta;
RA   Strausberg R.;
RL   Submitted (MAR-2004) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC067824; AAH67824.1; -.
DR   HSSP; Q07960; 1AM4.
DR   InterPro; IPR001251; CRAL_TRIO_C.
DR   InterPro; IPR000198; RhoGAP.
DR   InterPro; IPR008936; Rho_GAP.
DR   SMART; SM00324; RhoGAP; 1.
DR   SMART; SM00516; SEC14; 1.
DR   PROSITE; PS50191; CRAL_TRIO; 1.
DR   PROSITE; PS50238; RHOGAP; 1.
FT   NON_TER       1      1
SQ   SEQUENCE   337 AA;  38956 MW;  DB3921FA61C78C92 CRC64;

  Query Match             30.1%;  Score 55;  DB 2;  Length 337;
  Best Local Similarity   52.2%;  Pred. No. 1.1e+02;
  Matches   12;  Conservative    4;  Mismatches    7;  Indels    0;  Gaps    0;

Qy          2 GLSQEQLEHRERSLQTLRDIQRM 24
              ||  | |  |  |:||:|:|||:
Db        256 GLRTEGLFRRSASVQTVREIQRL 278


RESULT 31
HEM1_SYNEL
ID   HEM1_SYNEL     STANDARD;      PRT;   426 AA.
AC   Q8DI53;
DT   10-OCT-2003 (Rel. 42, Created)
DT   10-OCT-2003 (Rel. 42, Last sequence update)
DT   05-JUL-2004 (Rel. 44, Last annotation update)
DE   Glutamyl-tRNA reductase (EC 1.2.1.-) (GluTR).
GN   Name=hemA; OrderedLocusNames=tll1738;
OS   Synechococcus elongatus (Thermosynechococcus elongatus).
OC   Bacteria; Cyanobacteria; Chroococcales; Synechococcus.
OX   NCBI_TaxID=32046;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BP-1;
RX   MEDLINE=22225144; PubMed=12240834;
RA   Nakamura Y., Kaneko T., Sato S., Ikeuchi M., Katoh H., Sasamoto S.,
RA   Watanabe A., Iriguchi M., Kawashima K., Kimura T., Kishida Y.,
RA   Kiyokawa C., Kohara M., Matsumoto M., Matsuno A., Nakazaki N.,
RA   Shimpo S., Sugimoto M., Takeuchi C., Yamada M., Tabata S.;
RT   "Complete genome structure of the thermophilic cyanobacterium
RT   Thermosynechococcus elongatus BP-1.";
RL   DNA Res. 9:123-130(2002).
CC   -!- CATALYTIC ACTIVITY: Glutamyl-tRNA(Glu) + NADPH = glutamate-1-
CC       semialdehyde + NADP(+) + tRNA(Glu).
CC   -!- PATHWAY: Porphyrin biosynthesis by the C5 pathway; first step.
CC       Involved in chlorophyll biosynthesis.
CC   -!- SIMILARITY: Belongs to the glutamyl-tRNA reductase family.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AP005375; BAC09290.1; -.
DR   HSSP; Q9UXR8; 1GPJ.
DR   HAMAP; MF_00087; -; 1.
```

```
DR   InterPro; IPR000343; GlutR.
DR   InterPro; IPR000594; ThiF_domain.
DR   Pfam; PF00745; GlutR_dimer; 1.
DR   Pfam; PF05201; GlutR_N; 1.
DR   Pfam; PF05200; GlutR_NAD_bind; 1.
DR   TIGRFAMs; TIGR01035; hemA; 1.
DR   PROSITE; PS00747; GLUTR; FALSE_NEG.
KW   Chlorophyll biosynthesis; Complete proteome; NADP; Oxidoreductase;
KW   Porphyrin biosynthesis.
FT   ACT_SITE     50     50       Nucleophile (By similarity).
FT   ACT_SITE     99     99       Proton acceptor (By similarity).
SQ   SEQUENCE   426 AA;  47596 MW;  D84CE5A1D2AA777E CRC64;

  Query Match             30.1%;  Score 55;  DB 1;  Length 426;
  Best Local Similarity   55.6%;  Pred. No. 1.4e+02;
  Matches   10;  Conservative    5;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          4 SQEQLEHRERSLQTLRDI 21
              ||  || |:|::|||:|:
Db        401 SQRDLESRQRAMQTLQDL 418


RESULT 32
Q6PJW1
ID   Q6PJW1      PRELIMINARY;      PRT;   428 AA.
AC   Q6PJW1;
DT   05-JUL-2004 (TrEMBLrel. 27, Created)
DT   05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT   05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE   Hypothetical protein (Fragment).
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Placenta;
RX   MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA   Krzywinski M.I., Skalska U., Smailus D.E., Schnerch A., Schein J.E.,
RA   Jones S.J., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length human
RT   and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Placenta;
RA   Strausberg R.;
RL   Submitted (JUL-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC010490; AAH10490.1; -.
DR   HSSP; Q07960; 1AM4.
DR   InterPro; IPR001251; CRAL_TRIO_C.
DR   InterPro; IPR000198; RhoGAP.
DR   InterPro; IPR008936; Rho_GAP.
DR   Pfam; PF00620; RhoGAP; 1.
DR   SMART; SM00324; RhoGAP; 1.
DR   SMART; SM00516; SEC14; 1.
DR   PROSITE; PS50191; CRAL_TRIO; 1.
DR   PROSITE; PS50238; RHOGAP; 1.
KW   Hypothetical protein.
FT   NON_TER       1      1
SQ   SEQUENCE   428 AA;  48607 MW;  5DE5828FF2043024 CRC64;

  Query Match             30.1%;  Score 55;  DB 2;  Length 428;
  Best Local Similarity   52.2%;  Pred. No. 1.4e+02;
  Matches   12;  Conservative    4;  Mismatches    7;  Indels    0;  Gaps    0;

Qy          2 GLSQEQLEHRERSLQTLRDIQRM 24
              || | |  |  |:||:|:|||:
Db        219 GLRTEGLFRRSASVQTVREIQRL 241


RESULT 33
Q8IZM6
ID   Q8IZM6      PRELIMINARY;      PRT;   433 AA.
AC   Q8IZM6;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   BCH domain-containing Cdc42GAP-like protein.
GN   Name=BPGAP1;
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=22970011; PubMed=12944407; DOI=10.1074/jbc.M304514200;
RA   Shang X., Zhou Y.T., Low B.C.;
RT   "Concerted regulation of cell dynamics by BNIP-2 and Cdc42GAP
RT   homology/Sec14p-like, proline-rich, and GTPase-activating protein
RT   domains of a novel Rho GTPase-activating protein, BPGAP1.";
RL   J. Biol. Chem. 278:45903-45914(2003).
DR   EMBL; AF544240; AAN40769.1; -.
DR   HSSP; Q07960; 1RGP.
DR   InterPro; IPR001251; CRAL_TRIO_C.
DR   InterPro; IPR000198; RhoGAP.
DR   InterPro; IPR008936; Rho_GAP.
DR   Pfam; PF00620; RhoGAP; 1.
DR   SMART; SM00324; RhoGAP; 1.
DR   SMART; SM00516; SEC14; 1.
DR   PROSITE; PS50191; CRAL_TRIO; 1.
DR   PROSITE; PS50238; RHOGAP; 1.
SQ   SEQUENCE   433 AA;  49691 MW;  0AE4B42A404AE1D3 CRC64;

  Query Match             30.1%;  Score 55;  DB 2;  Length 433;
  Best Local Similarity   52.2%;  Pred. No. 1.4e+02;
  Matches   12;  Conservative    4;  Mismatches    7;  Indels    0;  Gaps    0;

Qy          2 GLSQEQLEHRERSLQTLRDIQRM 24
              || | |  |  |:||:|:|||:
Db        224 GLRTEGLFRRSASVQTVREIQRL 246


RESULT 34
Q86XV6
ID   Q86XV6      PRELIMINARY;      PRT;   464 AA.
AC   Q86XV6;
DT   01-JUN-2003 (TrEMBLrel. 24, Created)
DT   01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   ARHGAP8 protein (Fragment).
GN   Name=ARHGAP8;
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Colon;
RX   MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
```

```
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA   Krzywinski M.I., Skalska U., Smailus D.E., Schnerch A., Schein J.E.,
RA   Jones S.J., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length human
RT   and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Colon;
RA   Strausberg R.;
RL   Submitted (MAR-2003) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC048280; AAH48280.1; -.
DR   HSSP; Q07960; 1RGP.
DR   InterPro; IPR001251; CRAL_TRIO_C.
DR   InterPro; IPR000198; RhoGAP.
DR   InterPro; IPR008936; Rho_GAP.
DR   Pfam; PF00620; RhoGAP; 1.
DR   SMART; SM00324; RhoGAP; 1.
DR   SMART; SM00516; SEC14; 1.
DR   PROSITE; PS50191; CRAL_TRIO; 1.
DR   PROSITE; PS50238; RHOGAP; 1.
FT   NON_TER       1      1
SQ   SEQUENCE   464 AA;  53142 MW;  82F68ACD8AD219C7 CRC64;

  Query Match             30.1%;  Score 55;  DB 2;  Length 464;
  Best Local Similarity   52.2%;  Pred. No. 1.5e+02;
  Matches   12;  Conservative    4;  Mismatches    7;  Indels    0;  Gaps    0;

Qy          2 GLSQEQLEHRERSLQTLRDIQRM 24
              ||  | |  |  |:||:|:|||:
Db        255 GLRTEGLFRRSASVQTVREIQRL 277


RESULT 35
RHG8_HUMAN
ID   RHG8_HUMAN     STANDARD;      PRT;   718 AA.
AC   Q9NSG0; O75983; O95695; Q96RW1; Q96RW2; Q9HA49; Q9HC46; Q9NVX8;
AC   Q9NXL1; Q9UH20;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   05-JUL-2004 (Rel. 44, Last annotation update)
DE   Rho-GTPase-activating protein 8 (PP610).
GN   Name=ARHGAP8;
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A. (ISOFORM 4).
RA   Gu J.R., Wan D.F., Zhao X.T., Zhou X.M., Jiang H.Q., Zhang P.P.,
RA   Qin W.X., Huang Y., Qiu X.K., Qian L.F., He L.P., Li H.N., Yu Y.,
RA   Yu J., Han L.H.;
RT   "Novel human cDNA clones with function of inhibiting cancer cell
RT   growth.";
RL   Submitted (AUG-1999) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).
RA   McDermid H.E., Hu S., Grundy P., Trichet V.;
RT   "ARHGAP8: a putative tumor-suppressor gene on chromosome 22q13.3.";
RL   Submitted (OCT-1999) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A. (ISOFORMS 5; 6 AND 7).
RC   TISSUE=Colon mucosa, and Mammary gland;
RX   PubMed=14702039; DOI=10.1038/ng1285;
RA   Ota T., Suzuki Y., Nishikawa T., Otsuki T., Sugiyama T., Irie R.,
RA   Wakamatsu A., Hayashi K., Sato H., Nagai K., Kimura K., Makita H.,
RA   Sekine M., Obayashi M., Nishi T., Shibahara T., Tanaka T., Ishii S.,
RA   Yamamoto J.-I., Saito K., Kawai Y., Isono Y., Nakamura Y.,
RA   Nagahari K., Murakami K., Yasuda T., Iwayanagi T., Wagatsuma M.,
RA   Shiratori A., Sudo H., Hosoiri T., Kaku Y., Kodaira H., Kondo H.,
RA   Sugawara M., Takahashi M., Kanda K., Yokoi T., Furuya T., Kikkawa E.,
RA   Omura Y., Abe K., Kamihara K., Katsuta N., Sato K., Tanikawa M.,
RA   Yamazaki M., Ninomiya K., Ishibashi T., Yamashita H., Murakawa K.,
RA   Fujimori K., Tanai H., Kimata M., Watanabe M., Hiraoka S., Chiba Y.,
RA   Ishida S., Ono Y., Takiguchi S., Watanabe S., Yosida M., Hotuta T.,
RA   Kusano J., Kanehori K., Takahashi-Fujii A., Hara H., Tanase T.-O.,
RA   Nomura Y., Togiya S., Komai F., Hara R., Takeuchi K., Arita M.,
RA   Imose N., Musashino K., Yuuki H., Oshima A., Sasaki N., Aotsuka S.,
RA   Yoshikawa Y., Matsunawa H., Ichihara T., Shiohata N., Sano S.,
RA   Moriya S., Momiyama H., Satoh N., Takami S., Terashima Y., Suzuki O.,
RA   Nakagawa S., Senoh A., Mizoguchi H., Goto Y., Shimizu F., Wakebe H.,
RA   Hishigaki H., Watanabe T., Sugiyama A., Takemoto M., Kawakami B.,
RA   Yamazaki M., Watanabe K., Kumagai A., Itakura S., Fukuzumi Y.,
RA   Fujimori Y., Komiyama M., Tashiro H., Tanigami A., Fujiwara T.,
RA   Ono T., Yamada K., Fujii Y., Ozaki K., Hirao M., Ohmori Y.,
RA   Kawabata A., Hikiji T., Kobatake N., Inagaki H., Ikema Y., Okamoto S.,
RA   Okitani R., Kawakami T., Noguchi S., Itoh T., Shigeta K., Senba T.,
RA   Matsumura K., Nakajima Y., Mizuno T., Morinaga M., Sasaki M.,
RA   Togashi T., Oyama M., Hata H., Watanabe M., Komatsu T.,
RA   Mizushima-Sugano J., Satoh T., Shirai Y., Takahashi Y., Nakagawa K.,
RA   Okumura K., Nagase T., Nomura N., Kikuchi H., Masuho Y., Yamashita R.,
RA   Nakai K., Yada T., Nakamura Y., Ohara O., Isogai T., Sugano S.;
RT   "Complete sequencing and characterization of 21,243 full-length human
RT   cDNAs.";
RL   Nat. Genet. 36:40-45(2004).
RN   [4]
RP   SEQUENCE FROM N.A. (ISOFORM 3).
RA   Goward M.E., Huckle E.J.;
RL   Submitted (APR-2000) to the EMBL/GenBank/DDBJ databases.
RN   [5]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20057165; PubMed=10591208; DOI=10.1038/990031;
RA   Dunham I., Hunt A.R., Collins J.E., Bruskiewich R., Beare D.M.,
RA   Clamp M., Smink L.J., Ainscough R., Almeida J.P., Babbage A.K.,
RA   Bagguley C., Bailey J., Barlow K.F., Bates K.N., Beasley O.P.,
RA   Bird C.P., Blakey S.E., Bridgeman A.M., Buck D., Burgess J.,
RA   Burrill W.D., Burton J., Carder C., Carter N.P., Chen Y., Clark G.,
RA   Clegg S.M., Cobley V.E., Cole C.G., Collier R.E., Connor R.,
RA   Conroy D., Corby N.R., Coville G.J., Cox A.V., Davis J., Dawson E.,
RA   Dhami P.D., Dockree C., Dodsworth S.J., Durbin R.M., Ellington A.G.,
RA   Evans K.L., Fey J.M., Fleming K., French L., Garner A.A.,
RA   Gilbert J.G.R., Goward M.E., Grafham D.V., Griffiths M.N.D., Hall C.,
RA   Hall R.E., Hall-Tamlyn G., Heathcott R.W., Ho S., Holmes S.,
RA   Hunt S.E., Jones M.C., Kershaw J., Kimberley A.M., King A.,
RA   Laird G.K., Langford C.F., Leversha M.A., Lloyd C., Lloyd D.M.,
RA   Martyn I.D., Mashreghi-Mohammadi M., Matthews L.H., Mccann O.T.,
RA   Mcclay J., Mclaren S., McMurray A.A., Milne S.A., Mortimore B.J.,
RA   Odell C.N., Pavitt R., Pearce A.V., Pearson D., Phillimore B.J.C.T.,
RA   Phillips S.H., Plumb R.W., Ramsay H., Ramsey Y., Rogers L., Ross M.T.,
RA   Scott C.E., Sehra H.K., Skuce C.D., Smalley S., Smith M.L.,
RA   Soderlund C., Spragon L., Steward C.A., Sulston J.E., Swann R.M.,
RA   Vaudin M., Wall M., Wallis J.M., Whiteley M.N., Willey D.L.,
RA   Williams L., Williams S.A., Williamson H., Wilmer T.E., Wilming L.,
RA   Wright C.L., Hubbard T., Bentley D.R., Beck S., Rogers J., Shimizu N.,
RA   Minoshima S., Kawasaki K., Sasaki T., Asakawa S., Kudoh J.,
RA   Shintani A., Shibuya K., Yoshizaki Y., Aoki N., Mitsuyama S.,
RA   Roe B.A., Chen F., Chu L., Crabtree J., Deschamps S., Do A., Do T.,
RA   Dorman A., Fang F., Fu Y., Hu P., Hua A., Kenton S., Lai H., Lao H.I.,
RA   Lewis J., Lewis S., Lin S.-P., Loh P., Malaj E., Nguyen T., Pan H.,
RA   Phan S., Qi S., Qian Y., Ray L., Ren Q., Shaull S., Sloan D., Song L.,
RA   Wang Q., Wang Y., Wang Z., White J., Willingham D., Wu H., Yao Z.,
RA   Zhan M., Zhang G., Chissoe S., Murray J., Miller N., Minx P.,
RA   Fulton R., Johnson D., Bemis G., Bentley D., Bradshaw H., Bourne S.,
```

```
RA   Cordes M., Du Z., Fulton L., Goela D., Graves T., Hawkins J.,
RA   Hinds K., Kemp K., Latreille P., Layman D., Ozersky P., Rohlfing T.,
RA   Scheet P., Walker C., Wamsley A., Wohldmann P., Pepin K., Nelson J.,
RA   Korf I., Bedell J.A., Hillier L.W., Mardis E., Waterston R.,
RA   Wilson R., Emanuel B.S., Shaikh T., Kurahashi H., Saitta S.,
RA   Budarf M.L., McDermid H.E., Johnson A., Wong A.C.C., Morrow B.E.,
RA   Edelmann L., Kim U.J., Shizuya H., Simon M.I., Dumanski J.P.,
RA   Peyrard M., Kedra D., Seroussi E., Fransson I., Tapia I., Bruder C.E.,
RA   O'Brien K.P., Wilkinson P., Bodenteich A., Hartman K., Hu X.,
RA   Khan A.S., Lane L., Tilahun Y., Wright H.;
RT   "The DNA sequence of human chromosome 22.";
RL   Nature 402:489-495(1999).
CC   -!- FUNCTION: GTPase activator for the Rho-type GTPases by converting
CC       them to an inactive GDP-bound state (By similarity).
CC   -!- ALTERNATIVE PRODUCTS:
CC       Event=Alternative splicing; Named isoforms=8;
CC         Comment=Additional isoforms seem to exist. Full isoforms so far
CC         detected are isoforms 1 to 7. Experimental confirmation may be
CC         lacking for some isoforms;
CC       Name=8;
CC         IsoId=Q9NSG0-1; Sequence=Displayed;
CC       Name=1;
CC         IsoId=Q9NSG0-2; Sequence=VSP_001645, VSP_001649, VSP_001652;
CC       Name=2;
CC         IsoId=Q9NSG0-3; Sequence=VSP_001645, VSP_001649, VSP_001653,
CC                                  VSP_001655;
CC       Name=3;
CC         IsoId=Q9NSG0-4; Sequence=VSP_001651;
CC       Name=4;
CC         IsoId=Q9NSG0-5; Sequence=VSP_001650, VSP_001654;
CC       Name=5;
CC         IsoId=Q9NSG0-6; Sequence=VSP_001647;
CC       Name=6;
CC         IsoId=Q9NSG0-7; Sequence=VSP_001647, VSP_001652, VSP_001656,
CC                                  VSP_001657;
CC       Name=7;
CC         IsoId=Q9NSG0-8; Sequence=VSP_001646, VSP_001648, VSP_001650,
CC                                  VSP_001654;
CC   -!- SIMILARITY: Contains 1 CRAL-TRIO domain.
CC   -!- SIMILARITY: Contains 1 Rho-GAP domain.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF177331; AAG17975.1; -.
DR   EMBL; AF195968; AAK58136.1; -.
DR   EMBL; AF195969; AAK58137.1; -.
DR   EMBL; AK000192; BAA90999.1; -.
DR   EMBL; AK001306; BAA91614.1; -.
DR   EMBL; AK022305; BAB14008.1; -.
DR   EMBL; AL355192; CAB90248.1; -.
DR   EMBL; Z98743; CAB11416.1; ALT_INIT.
DR   EMBL; Z93244; -; NOT_ANNOTATED_CDS.
DR   EMBL; Z83838; CAB62993.1; ALT_INIT.
DR   PIR; B59436; B59436.
DR   HSSP; Q07960; 1RGP.
DR   Genew; HGNC:677; ARHGAP8.
DR   InterPro; IPR001251; CRAL_TRIO_C.
DR   InterPro; IPR008936; Rho_GAP.
DR   InterPro; IPR000198; RhoGAP.
DR   Pfam; PF00620; RhoGAP; 1.
DR   PROSITE; PS50191; CRAL_TRIO; 1.
DR   PROSITE; PS50238; RHOGAP; 1.
KW   Alternative splicing; GTPase activation.
FT   DOMAIN      267    453       CRAL-TRIO.
FT   DOMAIN      480    666       Rho-GAP.
FT   VARSPLIC      1     95       Missing (in isoform 1 and isoform 2).
FT                                /FTId=VSP_001645.
FT   VARSPLIC      1    101       Missing (in isoform 7).
FT                                /FTId=VSP_001646.
FT   VARSPLIC      1    254       Missing (in isoform 5 and isoform 6).
FT                                /FTId=VSP_001647.
FT   VARSPLIC    102    107       KIRFYE -> MAPMPT (in isoform 7).
FT                                /FTId=VSP_001648.
FT   VARSPLIC    108    230       Missing (in isoform 1 and isoform 2).
FT                                /FTId=VSP_001649.
FT   VARSPLIC    232    388       LQRDKAAAAAVLGAVRKRPSVVPMAGQDPALSTSHPFYDVA
FT                                RHGILQVAGDDRFGRRVVTFSCCRMPPSHELDHQRLLEYLK
FT                                YTLDQYVENDYTIVYFHYGLNSRNKPSLGWLQSAYKEPDRK
FT                                DGDLTMWPRLVSNSKLKRSSHLSLPKYWDYRYKK -> KRL
FT                                LRRSRSGDVLAKNPVVRSKSYNTPLLNPVQEHEAEGAAAGG
FT                                TSIRRHSVSEMTSCPEPQGFSDPPGQGPTGTFRSSPAPHSG
FT                                PCPSRLYPTTQPPEQGLDPTRSSLPRSSPENLVDQILESVD
FT                                SDSEGIFIDFGRGRGSGMSDLEGSGGRQSVV (in
FT                                isoform 4 and isoform 7).
FT                                /FTId=VSP_001650.
FT   VARSPLIC    311    385       Missing (in isoform 3).
FT                                /FTId=VSP_001651.
FT   VARSPLIC    355    385       Missing (in isoform 1 and isoform 6).
FT                                /FTId=VSP_001652.
FT   VARSPLIC    386    451       YKKNLKALYVVHPTSFIKVLWNILKPLISHKFGKKVIYFNY
FT                                LSELHEHLKYDQLVIPPEVLRYDEK -> QEPPCQANTLVL
FT                                KGPDSQHFSFAGLLLYCNNSAGLCSSKTLWTLKCEFHVIFT
FT                                CHEIFFCFFFSTT (in isoform 2).
FT                                /FTId=VSP_001653.
FT   VARSPLIC    389    718       Missing (in isoform 4 and isoform 7).
FT                                /FTId=VSP_001654.

  Query Match             30.1%;  Score 55;  DB 1;  Length 718;
  Best Local Similarity   52.2%;  Pred. No. 2.5e+02;
  Matches   12;  Conservative    4;  Mismatches    7;  Indels    0;  Gaps    0;

Qy          2 GLSQEQLEHRERSLQTLRDIQRM 24
              ||  | |  |  |:||:|:|||:
Db        509 GLRTEGLFRRSASVQTVREIQRL 531


RESULT 36
ASC3_MOUSE
ID   ASC3_MOUSE     STANDARD;      PRT;   174 AA.
AC   Q9JJR7;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   05-JUL-2004 (Rel. 44, Last annotation update)
DE   Achaete-scute homolog 3 (bHLH transcriptional regulator Sgn-1) (Mash-
DE   3).
GN   Name=Ascl3; Synonyms=Mash3, Sgn1;
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Kemp P.R., Cooper W.N., Metcalfe J.C.;
RT   "MASH3 a novel basic helix-loop-helix protein that inhibits myogenesis
RT   in C2C12 cells.";
RL   Submitted (APR-2000) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6;
RX   MEDLINE=21643927; PubMed=11784080; DOI=10.1006/dbio.2001.0473;
RA   Yoshida S., Ohbo K., Takakura A., Takebayashi H., Okada T., Abe K.,
RA   Nabeshima Y.;
RT   "Sgn1, a basic helix-loop-helix transcription factor delineates the
RT   salivary gland duct cell lineage in mice.";
RL   Dev. Biol. 240:517-530(2001).
RN   [3]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=21418998; PubMed=11528127;
RA   Amid C., Bahr A., Mujica A., Sampson N., Bikar S.E., Winterpacht A.,
RA   Zabel B., Hankeln T., Schmidt E.R.;
```

```
RT   "Comparative genomic sequencing reveals a strikingly similar
RT   architecture of a conserved syntenic region on human chromosome
RT   11p15.3 (including gene ST5) and mouse chromosome 7.";
RL   Cytogenet. Cell Genet. 93:284-290(2001).
CC   -!- FUNCTION: Transcriptional repressor. Inhibits myogenesis.
CC   -!- SUBUNIT: Efficient DNA binding requires dimerization with another
CC       bHLH protein.
CC   -!- SUBCELLULAR LOCATION: Nuclear.
CC   -!- TISSUE SPECIFICITY: Specifically expressed in the salivary duct
CC       cells.
CC   -!- SIMILARITY: Contains 1 basic helix-loop-helix (bHLH) domain.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AJ277605; CAC37689.1; -.
DR   EMBL; AB046448; BAB83911.1; -.
DR   EMBL; AJ400878; CAB92296.1; -.
DR   MGD; MGI:1928820; Ascl3.
DR   GO; GO:0005634; C:nucleus; IDA.
DR   GO; GO:0005667; C:transcription factor complex; IPI.
DR   GO; GO:0003677; F:DNA binding; IDA.
DR   GO; GO:0005515; F:protein binding; IPI.
DR   GO; GO:0030528; F:transcription regulator activity; IDA.
DR   GO; GO:0006357; P:regulation of transcription from Pol II pro. . .; IDA.
DR   InterPro; IPR001092; HLH_basic.
DR   Pfam; PF00010; HLH; 1.
DR   SMART; SM00353; HLH; 1.
DR   PROSITE; PS50888; HLH; 1.
KW   DNA-binding; Nuclear protein; Repressor; Transcription regulation.
FT   DNA_BIND     95    105       Basic motif.
FT   DOMAIN      106    145       Helix-loop-helix motif.
SQ   SEQUENCE   174 AA;  20245 MW;  D89E56C8A9D3440B CRC64;

  Query Match             29.8%;  Score 54.5;  DB 1;  Length 174;
  Best Local Similarity   36.1%;  Pred. No. 59;
  Matches   13;  Conservative    7;  Mismatches    9;  Indels    7;  Gaps    1;

Qy          3 LSQEQLEHRERSLQTLRD-------IQRMLFPDEKE 31
              | :: || |   ::|||        :| :|:||| |
Db        119 LPEDYLEKRLSKVETLRAAIKYISYLQSLLYPDESE 154


RESULT 37
PHYB_ORYSA
ID   PHYB_ORYSA     STANDARD;      PRT;  1171 AA.
AC   P25764;
DT   01-MAY-1992 (Rel. 22, Created)
DT   01-MAY-1992 (Rel. 22, Last sequence update)
DT   05-JUL-2004 (Rel. 44, Last annotation update)
DE   Phytochrome B.
GN   Name=PHYB; Synonyms=PHYB1;
OS   Oryza sativa (Rice).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC   Ehrhartoideae; Oryzeae; Oryza.
OX   NCBI_TaxID=4530;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=cv. Indica / IR36; TISSUE=Seedling shoot;
RX   MEDLINE=91172131; PubMed=2005872;
RA   Dehesh K., Tepperman J., Christensen A.H., Quail P.H.;
RT   "phyB is evolutionarily conserved and constitutively expressed in rice
RT   seedling shoots.";
RL   Mol. Gen. Genet. 225:305-313(1991).
CC   -!- FUNCTION: Regulatory photoreceptor which exists in two forms that
CC       are reversibly interconvertible by light: the Pr form that absorbs
CC       maximally in the red region of the spectrum and the Pfr form that
CC       absorbs maximally in the far-red region. Photoconversion of Pr in
CC       Pfr induces an array of morphogenic responses, whereas
CC       reconversion of Pfr to Pr cancels the induction of those
CC       responses. Pfr controls the expression of a number of nuclear
CC       genes including those encoding the small subunit of ribulose-
CC       bisphosphate carboxylase, chlorophyll A/B binding protein,
CC       protochlorophyllide reductase, rRNA, etc. It also controls the
CC       expression of its own gene(s) in a negative feedback fashion.
CC   -!- SUBUNIT: Homodimer.
CC   -!- PTM: Contains one covalently linked tetrapyrrole chromophore.
CC   -!- SIMILARITY: Belongs to the phytochrome family.
CC   -!- SIMILARITY: Contains 1 histidine kinase domain.
CC   -!- SIMILARITY: Contains 2 PAS (PER-ARNT-SIM) dimerization domains.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X57563; CAA40795.2; -.
DR   PIR; S14065; S14065.
DR   Gramene; P25764; -.
DR   InterPro; IPR003594; ATPbind_ATPase.
DR   InterPro; IPR003018; GAF.
DR   InterPro; IPR005467; His_kinase.
DR   InterPro; IPR003661; His_kinA_N.
DR   InterPro; IPR000014; PAS.
DR   InterPro; IPR001294; Phytochrome.
DR   Pfam; PF01590; GAF; 1.
DR   Pfam; PF02518; HATPase_c; 1.
DR   Pfam; PF00512; HisKA; 1.
DR   Pfam; PF00989; PAS; 2.
DR   Pfam; PF00360; Phytochrome; 1.
DR   PRINTS; PR01033; PHYTOCHROME.
DR   SMART; SM00065; GAF; 1.
DR   SMART; SM00387; HATPase_c; 1.
DR   SMART; SM00388; HisKA; 1.
DR   SMART; SM00091; PAS; 2.
DR   TIGRFAMs; TIGR00229; sensory_box; 2.
DR   PROSITE; PS50109; HIS_KIN; 1.
DR   PROSITE; PS50112; PAS; 2.
DR   PROSITE; PS00245; PHYTOCHROME_1; 1.
DR   PROSITE; PS50046; PHYTOCHROME_2; 1.
KW   Chromophore; Multigene family; Photoreceptor; Phytochrome; Repeat;
KW   Transcription regulation.
FT   DOMAIN      661    732       PAS 1.
FT   DOMAIN      795    866       PAS 2.
FT   DOMAIN      943   1161       Histidine kinase.
FT   DOMAIN       39     51       Poly-Gly.
FT   BINDING     364    364       Chromophore (By similarity).
SQ   SEQUENCE   1171 AA;  128384 MW;  E8292E88B769BF16 CRC64;

  Query Match             29.8%;  Score 54.5;  DB 1;  Length 1171;
  Best Local Similarity   48.1%;  Pred. No. 5e+02;
  Matches   13;  Conservative    4;  Mismatches    5;  Indels    5;  Gaps    1;

Qy          3 LSQEQLEHRERSLQTLRDIQRMLFPDE 29
              :||  :: ||| || :|||     |||
Db       1019 VSQVMIQLRERDLQLIRDI-----PDE 1040


RESULT 38
Q84LN8
ID   Q84LN8      PRELIMINARY;      PRT;  1171 AA.
AC   Q84LN8;
DT   01-JUN-2003 (TrEMBLrel. 24, Created)
DT   01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT   01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE   Phytochrome B.
GN   Name=PHYB;
```

```
OS   Oryza sativa (japonica cultivar-group).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC   Ehrhartoideae; Oryzeae; Oryza.
OX   NCBI_TaxID=39947;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Tahir M., Kanegae H., Takano M.;
RL   Submitted (MAY-2003) to the EMBL/GenBank/DDBJ databases.
CC   -!- SIMILARITY: Contains 1 histidine kinase domain.
DR   EMBL; AB109892; BAC76432.1; -.
DR   Gramene; Q84LN8; -.
DR   GO; GO:0016020; C:membrane; IEA.
DR   GO; GO:0005524; F:ATP binding; IEA.
DR   GO; GO:0008020; F:G-protein coupled photoreceptor activity; IEA.
DR   GO; GO:0016301; F:kinase activity; IEA.
DR   GO; GO:0000155; F:two-component sensor molecule activity; IEA.
DR   GO; GO:0018298; P:protein-chromophore linkage; IEA.
DR   GO; GO:0009585; P:red, far-red light phototransduction; IEA.
DR   GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR   GO; GO:0007165; P:signal transduction; IEA.
DR   GO; GO:0007601; P:visual perception; IEA.
DR   InterPro; IPR003594; ATPbind_ATPase.
DR   InterPro; IPR000977; DNA_ligase.
DR   InterPro; IPR003018; GAF.
DR   InterPro; IPR005467; His_kinase.
DR   InterPro; IPR003661; His_kinA_N.
DR   InterPro; IPR000014; PAS.
DR   InterPro; IPR001294; Phytochrome.
DR   InterPro; IPR001680; WD40.
DR   Pfam; PF01590; GAF; 1.
DR   Pfam; PF02518; HATPase_c; 1.
DR   Pfam; PF00512; HisKA; 1.
DR   Pfam; PF00989; PAS; 2.
DR   Pfam; PF00360; Phytochrome; 1.
DR   PRINTS; PR01033; PHYTOCHROME.
DR   SMART; SM00065; GAF; 1.
DR   SMART; SM00387; HATPase_c; 1.
DR   SMART; SM00388; HisKA; 1.
DR   SMART; SM00091; PAS; 2.
DR   TIGRFAMs; TIGR00229; sensory_box; 1.
DR   PROSITE; PS00697; DNA_LIGASE_A1; UNKNOWN_1.
DR   PROSITE; PS50109; HIS_KIN; 1.
DR   PROSITE; PS50112; PAS; 2.
DR   PROSITE; PS00245; PHYTOCHROME_1; 1.
DR   PROSITE; PS50046; PHYTOCHROME_2; 1.
DR   PROSITE; PS00678; WD_REPEATS_1; UNKNOWN_1.
KW   Chromophore; Photoreceptor; Phytochrome.
SQ   SEQUENCE   1171 AA;  128492 MW;  DEE981FC89D46FDC CRC64;

  Query Match             29.8%;  Score 54.5;  DB 2;  Length 1171;
  Best Local Similarity   48.1%;  Pred. No. 5e+02;
  Matches   13;  Conservative    4;  Mismatches    5;  Indels    5;  Gaps    1;

Qy          3 LSQEQLEHRERSLQTLRDIQRMLFPDE 29
              :||  :: ||| || :|||     |||
Db       1019 VSQVMIQLRERDLQLIRDI-----PDE 1040


RESULT 39
Q6D1R6
ID   Q6D1R6      PRELIMINARY;      PRT;   326 AA.
AC   Q6D1R6;
DT   25-OCT-2004 (TrEMBLrel. 28, Created)
DT   25-OCT-2004 (TrEMBLrel. 28, Last sequence update)
DT   25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DE   Pyoverdine biosynthesis protein.
GN   Name=pvcA; OrderedLocusNames=ECA3381;
OS   Erwinia carotovora (subsp. atroseptica) (Pectobacterium atrosepticum).
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC   Enterobacteriaceae; Pectobacterium.
OX   NCBI_TaxID=29471;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=SCRI 1043 / ATCC BAA-672;
RX   PubMed=15263089; DOI=10.1073/pnas.0402424101;
RA   Bell K.S., Sebaihia M., Pritchard L., Holden M.T.G., Hyman L.J.,
RA   Holeva M.C., Thomson N.R., Bentley S.D., Churcher L.J.C., Mungall K.,
RA   Atkin R., Bason N., Brooks K., Chillingworth T., Clark K., Doggett J.,
RA   Fraser A., Hance Z., Hauser H., Jagels K., Moule S., Norbertczak H.,
RA   Ormond D., Price C., Quail M.A., Sanders M., Walker D., Whitehead S.,
RA   Salmond G.P.C., Birch P.R.J., Parkhill J., Toth I.K.;
RT   "Genome sequence of the enterobacterial phytopathogen Erwinia
RT   carotovora subsp. atroseptica and characterization of virulence
RT   factors.";
RL   Proc. Natl. Acad. Sci. U.S.A. 101:11105-11110(2004).
DR   EMBL; BX950851; CAG76279.1; -.
DR   InterPro; IPR007817; DIT1_PvcA.
DR   Pfam; PF05141; DIT1_PvcA; 1.
KW   Complete proteome.
SQ   SEQUENCE   326 AA;  37006 MW;  FAE727C53CC6B78B CRC64;

  Query Match             29.5%;  Score 54;  DB 2;  Length 326;
  Best Local Similarity   39.4%;  Pred. No. 1.4e+02;
  Matches   13;  Conservative    5;  Mismatches   13;  Indels    2;  Gaps    1;

Qy          5 QEQLEHRERSLQTLRDIQRMLFPD--EKEFTGA 35
              :|||   |  ||  | | | |: |    ::||:
Db        187 KEQLMQSEEGLQLYRSITRFLYEDSLRPDYTGS 219


RESULT 40
Q7R1P0
ID   Q7R1P0      PRELIMINARY;      PRT;   554 AA.
AC   Q7R1P0;
DT   01-MAR-2004 (TrEMBLrel. 26, Created)
DT   01-MAR-2004 (TrEMBLrel. 26, Last sequence update)
DT   01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE   GLP_28_36755_35091.
OS   Giardia lamblia ATCC 50803.
OC   Eukaryota; Diplomonadida; Hexamitidae; Giardiinae; Giardia.
OX   NCBI_TaxID=184922;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=WB C6;
RA   Morrison H.G., McArthur A.G., Adam R.D., Aley S.B., Gillin F.D.,
RA   Olsen G.J., Sogin M.L.;
RT   "Draft sequence of the Giardia lamblia genome.";
RL   Submitted (MAR-2003) to the EMBL/GenBank/DDBJ databases.
CC   -!- SIMILARITY: Belongs to the Ser/Thr protein kinase family.
CC   -!- CAUTION: The sequence shown here is derived from an
CC       EMBL/GenBank/DDBJ whole genome shotgun (WGS) entry which is
CC       preliminary data.
DR   EMBL; AACB01000026; EAA41219.1; -.
DR   HSSP; P24941; 1E1X.
DR   GO; GO:0005524; F:ATP binding; IEA.
DR   GO; GO:0004674; F:protein serine/threonine kinase activity; IEA.
DR   GO; GO:0016740; F:transferase activity; IEA.
DR   GO; GO:0006468; P:protein amino acid phosphorylation; IEA.
DR   InterPro; IPR011009; Kinase_like.
DR   InterPro; IPR000719; Prot_kinase.
DR   InterPro; IPR008271; Ser_thr_pkin_AS.
DR   ProDom; PD000001; Prot_kinase; 1.
DR   PROSITE; PS00107; PROTEIN_KINASE_ATP; UNKNOWN_1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR   PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
KW   ATP-binding; Kinase; Serine/threonine-protein kinase; Transferase.
SQ   SEQUENCE   554 AA;  62825 MW;  0A159E9518D36EF0 CRC64;

  Query Match             29.5%;  Score 54;  DB 2;  Length 554;
  Best Local Similarity   43.3%;  Pred. No. 2.5e+02;
  Matches   13;  Conservative    4;  Mismatches   11;  Indels    2;  Gaps    1;

Qy          1 DGLSQEQLEHRERSLQTLRDIQRMLFPDEK 30
              : |    || |:   |   || |::|||||
Db        265 EALGYPSLEERQGLSQ--NDIYRVIFPDEK 292
```

Search completed: June 8, 2005, 03:22:59
Job time : 167.75 secs

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: June 8, 2005, 02:59:37 ; Search time 133.875 Seconds
(without alignments)
104.003 Million cell updates/sec

Title: US-09-915-543-15_COPY_349_384
Perfect score: 183
Sequence: 1 DGLSQEQLEHRERSLQTLRDIQRMLFPDEKEFTGAQ 36

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 2105692 seqs, 386760381 residues

Total number of hits satisfying chosen parameters: 2105692

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_16Dec04:*
1: geneseqp1980s:*
2: geneseqp1990s:*
3: geneseqp2000s:*
4: geneseqp2001s:*
5: geneseqp2002s:*
6: geneseqp2003as:*
7: geneseqp2003bs:*
8: geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	183	100.0	1394	8	ADQ18945	Adq18945 Human sof
2	183	100.0	1426	5	AAB71229	Aab71229 Human leg
3	183	100.0	1426	7	ABW01534	Abw01534 Human lgs
4	183	100.0	1426	7	ADJ70152	Adj70152 Human hea
5	183	100.0	1426	8	ADJ71903	Adj71903 Human Lgs
6	183	100.0	1435	4	ABB11808	Abb11808 Human BCL
7	178	97.3	35	8	ADJ71893	Adj71893 Human Lgs
8	109	59.6	320	5	AAU78461	Aau78461 Mouse bet
9	109	59.6	1494	5	AAU78460	Aau78460 Mouse bet
10	107	58.5	738	5	AAU78463	Aau78463 Human bet
11	107	58.5	1115	5	AAB71230	Aab71230 Human leg
12	107	58.5	1115	7	ABW01535	Abw01535 Human lgs
13	107	58.5	1115	8	ADJ71905	Adj71905 Human Lgs
14	66	36.1	35	7	ABW01529	Abw01529 Drosophil
15	66	36.1	35	8	ADJ71892	Adj71892 Fruit fly
16	66	36.1	1429	4	ABB58779	Abb58779 Drosophil
17	66	36.1	1464	5	AAB71228	Aab71228 D. melano
18	66	36.1	1464	7	ABW01527	Abw01527 Drosophil
19	66	36.1	1464	8	ADJ71911	Adj71911 Fruit fly
20	61.5	33.6	1014	6	ABP98879	Abp98879 Human mol
21	57.5	31.4	603	6	ABJ25853	Abj25853 Aspergill
22	57.5	31.4	618	6	ABJ26453	Abj26453 Aspergill
23	57	31.1	411	6	ABU44941	Abu44941 Protein e
24	57	31.1	1034	6	ABU47461	Abu47461 Protein e
25	56	30.6	425	8	ADP04653	Adp04653 Sea squir
26	56	30.6	584	6	ABR53351	Abr53351 Protein s
27	56	30.6	584	7	ADK63408	Adk63408 Disease t
28	56	30.6	584	8	ADN19362	Adn19362 Bacterial
29	56	30.6	818	7	ADM26229	Adm26229 Hyperther
30	56	30.6	1132	8	ADL83239	Adl83239 Human PRO
31	56	30.6	1132	8	ADQ17519	Adq17519 Human sof
32	56	30.6	1294	4	ABB63502	Abb63502 Drosophil
33	55.5	30.3	757	6	ABU17570	Abu17570 Protein e
34	55	30.1	237	8	ADP56607	Adp56607 Human bre
35	55	30.1	248	8	ADP56608	Adp56608 Human bre
36	55	30.1	294	4	AAB95073	Aab95073 Human pro
37	55	30.1	294	6	ABR82444	Abr82444 Human ARP
38	55	30.1	294	8	ADQ74859	Adq74859 Human and
39	55	30.1	333	7	ADM83551	Adm83551 Human Rho
40	55	30.1	337	8	ADP56606	Adp56606 Human bre
41	55	30.1	357	8	ADP56605	Adp56605 Human bre
42	55	30.1	390	5	ABG96285	Abg96285 Human ova
43	55	30.1	433	4	AAB68522	Aab68522 Human GTP
44	55	30.1	433	4	AAG63851	Aag63851 Amino aci
45	55	30.1	433	4	AAG63852	Aag63852 Amino aci

ALIGNMENTS

RESULT 1
ADQ18945
ID ADQ18945 standard; protein; 1394 AA.
XX
AC ADQ18945;
XX
DT 26-AUG-2004 (first entry)
XX
DE Human soft tissue sarcoma-upregulated protein - SEQ ID 1764.
XX
KW soft tissue sarcoma; cytostatic; gene therapy; vaccine; screening; human.
XX
OS Homo sapiens.
XX
PN WO2004048938-A2.
XX
PD 10-JUN-2004.
XX
PF 26-NOV-2003; 2003WO-US038193.
XX
PR 26-NOV-2002; 2002US-0429739P.
XX
PA (PROT-) PROTEIN DESIGN LABS INC.
XX
PI Aziz N, Ginsburg WM, Zlotnik A;
XX
DR WPI; 2004-441208/41.
XX
PT Early detection of soft tissue sarcoma comprises determining expression
PT of a gene in a first soft tissue sample and a normal soft tissue sample
PT and comparing the gene expression, also useful in treating soft tissue
PT sarcoma.
XX
PS Example 2; SEQ ID NO 1764; 210pp; English.
XX
CC The invention relates to a novel method for detecting soft tissue sarcoma
CC which comprises obtaining a first soft tissue sample from an individual
CC and a normal soft tissue sample from the same or different individual,
CC determining the expression of a gene in both samples and comparing the
CC expression of the gene in both soft tissue samples, where a higher level
CC of protein expression in the first soft tissue sample indicates the
CC presence of soft tissue sarcoma. The method of the invention has
CC cytostatic applications and may be useful for detecting soft tissue
CC sarcoma, possibly via gene therapy or vaccine production. The nucleic
CC acid sequences may be useful in diagnostic and screening applications.
CC The current sequence is that of a human soft tissue sarcoma-upregulated
CC protein of the invention. The current sequence is not shown within the
CC specification per se but was submitted in CD format by the inventor.

```
XX
SQ   Sequence 1394 AA;

  Query Match             100.0%;  Score 183;  DB 8;  Length 1394;
  Best Local Similarity   100.0%;  Pred. No. 5.9e-16;
  Matches   36;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 DGLSQEQLEHRERSLQTLRDIQRMLFPDEKEFTGAQ 36
              ||||||||||||||||||||||||||||||||||||
Db        349 DGLSQEQLEHRERSLQTLRDIQRMLFPDEKEFTGAQ 384


RESULT 2
AAB71229
ID   AAB71229 standard; protein; 1426 AA.
XX
AC   AAB71229;
XX
DT   18-NOV-2002  (first entry)
XX
DE   Human legless homologue lgs/bc19 protein.
XX
KW   Legless; human; lgs; Wnt/Wingless signaling pathway; Wnt; Wg;
KW   tissue proliferation; tumour; cytostatic; cellular disorder; colon;
KW   blood disorder; cancer; breast; head and neck cancer; brain; thyroid;
KW   medulloblastoma; skin cancer; tissue regeneration; tissue repair.
XX
OS   Homo sapiens.
XX
PN   US2002086986-A1.
XX
PD   04-JUL-2002.
XX
PF   27-JUL-2001; 2001US-00915543.
XX
PR   28-JUL-2000; 2000US-0221502P.
XX
PA   (BASL/) BASLER K.
PA   (BRUN/) BRUNNER E.
PA   (FROE/) FROESCH B.
PA   (KRAM/) KRAMPS T.
PA   (PETE/) PETER O.
XX
PI   Basler K,  Brunner E,  Froesch B,  Kramps T,  Peter O;
XX
DR   WPI; 2002-635689/68.
DR   N-PSDB; AAF88467.
XX
PT   Novel polypeptide useful in therapeutic method for treating disorders of
PT   cell fate such as cell differentiation or cell proliferation.
XX
PS   Example II; Fig 8B; 41pp; English.
XX
CC   This invention describes a novel polypeptide sharing one or more
CC   homologous amino acid domains with the legless (lgs) protein, a
CC   downstream component of the Wnt/Wingless (Wnt/Wg) signaling pathway
CC   involved in the formation and maintenance of spatial arrangements and
CC   proliferation of tissues during development, and in the formation and
CC   growth of many human tumours. The products of the invention have
CC   cytostatic activity and can be used to treat cellular disorders, blood
CC   disorders and cancers caused by over-stimulation of the Wnt pathway,
CC   where the cancerous condition is colon, breast, head and neck, brain,
CC   thyroid, medulloblastoma or skin cancer. The product could also be used
CC   to promote tissue regeneration and repair. This sequence represents the
CC   human legless (lgs) protein homologue lgs/bc19 described in the
CC   disclosure of the invention
XX
SQ   Sequence 1426 AA;

  Query Match             100.0%;  Score 183;  DB 5;  Length 1426;
  Best Local Similarity   100.0%;  Pred. No. 6e-16;
  Matches   36;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 DGLSQEQLEHRERSLQTLRDIQRMLFPDEKEFTGAQ 36
              ||||||||||||||||||||||||||||||||||||
Db        349 DGLSQEQLEHRERSLQTLRDIQRMLFPDEKEFTGAQ 384


RESULT 3
ABW01534
ID   ABW01534 standard; protein; 1426 AA.
XX
AC   ABW01534;
XX
DT   15-JAN-2004  (first entry)
XX
DE   Human lgs/bc19 protein.
XX
KW   Legless protein; lgs; cell fate disorder; blood disease; gene therapy;
KW   cancer; tissue regeneration; tissue repair; cytostatic.
XX
OS   Homo sapiens.
XX
PN   US2003114413-A1.
XX
PD   19-JUN-2003.
XX
PF   19-DEC-2002; 2002US-00322579.
XX
PR   28-JUL-2000; 2000US-0221502P.
PR   27-JUL-2001; 2001US-00915543.
XX
PA   (UYZU-) UNIV ZURICH.
XX
PI   Basler K,  Brunner E,  Froesch B,  Kramps T,  Peter O;
XX
DR   WPI; 2003-829432/77.
DR   N-PSDB; AAD62642.
XX
PT   Novel Lgs polypeptide useful for isolation of Lgs-binding proteins,
PT   diagnosing disorders of cell fate, treating diseases such as cancer.
XX
PS   Example 2; Fig 8B; 0pp; English.
XX
CC   The invention relates to novel legless (lgs) proteins and polynucleotides
CC   encoding such proteins. Lgs sequences are useful for the treatment of
CC   disorders of cell fate such as differentiation or proliferation. The
CC   invention is used to treat blood disease or a cancerous condition
CC   characterised by over-stimulation of the Wnt pathway such as colon,
CC   breast, head and neck, brain, thyroid, medulloblastoma or skin cancer and
CC   is administered to prevent progression from a pre-neoplastic or non-
CC   malignant condition to a neoplastic or malignant state. It is
CC   administered to promote tissue regeneration and repair. The invention is
CC   also useful in the therapy of diseases cost by an over-activation of Wg
CC   pathway. It is useful for reducing lgs gene expression in an invertebrate
CC   or vertebrate organism or an invertebrate or vertebrate cell line. The
CC   invention is also useful in gene therapy. The present sequence is human
CC   lgs/bc19 protein used in the invention
XX
SQ   Sequence 1426 AA;

  Query Match             100.0%;  Score 183;  DB 7;  Length 1426;
  Best Local Similarity   100.0%;  Pred. No. 6e-16;
  Matches   36;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 DGLSQEQLEHRERSLQTLRDIQRMLFPDEKEFTGAQ 36
              ||||||||||||||||||||||||||||||||||||
Db        349 DGLSQEQLEHRERSLQTLRDIQRMLFPDEKEFTGAQ 384


RESULT 4
ADJ70152
ID   ADJ70152 standard; protein; 1426 AA.
XX
AC   ADJ70152;
XX
```

DT 06-MAY-2004 (first entry)
XX
DE Human heat mitochondrial protein as a therapeutic target SeqID1958.
XX
KW mitochondrial; human; screening assay; diabetes mellitus;
KW Huntington's disease; osteoarthritis;
KW Leber's hereditary optic neuropathy; LHON;
KW mitochondrial encephalopathy lactic acidosis and stroke; MELAS;
KW myoclonic epilepsy ragged red fibre syndrome; MERRF; cancer;
KW neuroprotective; nootropic; antidiabetic; anticonvulsant; antiarthritic;
KW osteopathic; ophthalmological; cytostatic.
XX
OS Homo sapiens.
XX
PN WO2003087768-A2.
XX
PD 23-OCT-2003.
XX
PF 04-APR-2003; 2003WO-US010870.
XX
PR 12-APR-2002; 2002US-0372843P.
PR 17-JUN-2002; 2002US-0389987P.
PR 20-SEP-2002; 2002US-0412418P.
XX
PA (MITO-) MITOKOR.
PA (BUCK-) BUCK INST AGE RES.
XX
PI Ghosh SS, Fahy ED, Zhang B, Gibson BW, Taylor SW, Glenn GM;
PI Warnock DE;
XX
DR WPI; 2003-845369/78.
XX
PT Identifying a mitochondrial target for drug screening assays and for
PT treating diseases associated with altered mitochondrial function,
PT comprises detecting a modified polypeptide in a sample and correlating
PT with the disease.
XX
PS Claim 1; SEQ ID NO 1958; 180pp; English.
XX
CC This invention relates to novel mitochondrial targets that can be used
CC for therapeutic intervention in treating a disease associated with
CC altered mitochondrial function. Specifically, it refers to a method for
CC identifying proteins of the human heart mitochondrial proteome that are
CC useful for drug screening assays, as well as therapeutic targets. The
CC present invention describes a method for identifying such proteins that
CC can be used in the treatment of various diseases associated with altered
CC mitochondrial function including diabetes mellitus, Huntington's disease,
CC osteoarthritis, Leber's hereditary optic neuropathy (LHON), mitochondrial
CC encephalopathy lactic acidosis and stroke (MELAS), myoclonic epilepsy
CC ragged red fibre syndrome (MERRF) or cancer. Accordingly, these
CC compositions have neuroprotective, nootropic, antidiabetic,
CC anticonvulsant, antiarthritic, osteopathic, ophthalmological and
CC cytostatic activities. This polypeptide sequence is a human heart
CC mitochondrial protein of the invention.
XX
SQ Sequence 1426 AA;

Query Match 100.0%; Score 183; DB 7; Length 1426;
Best Local Similarity 100.0%; Pred. No. 6e-16;
Matches 36; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 DGLSQEQLEHRERSLQTLRDIQRMLFPDEKEFTGAQ 36
||||||||||||||||||||||||||||||||||||
Db 349 DGLSQEQLEHRERSLQTLRDIQRMLFPDEKEFTGAQ 384

RESULT 5
ADJ71903
ID ADJ71903 standard; protein; 1426 AA.
XX
AC ADJ71903;
XX
DT 20-MAY-2004 (first entry)
XX
DE Human Lgs/Bc19 polypeptide.
XX
KW Human; legless; lgs; cell differentiation disorder;
KW cell proliferation disorder; cancer; Wnt pathway; medulloblastoma; colon;
KW breast; head; neck; brain; thyroid; skin; blood disease;
KW tissue regeneration; tissue repair; cytostatic; Lgs/Bc19.
XX
OS Homo sapiens.
XX
PN US2004038901-A1.
XX
PD 26-FEB-2004.
XX
PF 22-SEP-2003; 2003US-00664859.
XX
PR 28-JUL-2000; 2000US-0221502P.
PR 27-JUL-2001; 2001US-00915543.
XX
PA (UYZU-) UNIV ZURICH.
XX
PI Basler K, Brunner E, Froesch B, Kramps T, Peter O;
XX
DR WPI; 2004-203288/19.
DR N-PSDB; ADJ71902.
XX
PT Novel polypeptide sharing one or more homologue amino acid domains with
PT Legless protein being functional homologue of Legless, useful for
PT diagnosing disorders of cell fate.
XX
PS Example 2; SEQ ID NO 15; 62pp; English.
XX
CC The invention relates to a polypeptide sharing one or more homologous
CC amino acid domains with a Legless (Lgs) protein and is therefore a
CC functional homologue of Lgs. The invention also relates to a nucleotide
CC sequence encoding a protein present in invertebrate and/or vertebrate
CC organisms, the nucleotide sequence encoding a protein comprising a
CC positive function in a regulatory pathway and the use of the polypeptide
CC for the isolation of Lgs-binding proteins by carrying out an assay chosen
CC from an in vitro binding assay with such a peptide or a co-
CC immunoprecipitation from vertebrate or invertebrate cell lysates or a
CC mammalian or yeast two hybrid assay. The polypeptide and polynucleotide
CC are useful for treating disorders of cell fate, which involves
CC administering therapeutic compounds chosen from invertebrate and
CC vertebrate Lgs protein homologues or fragments, antibodies, antibody
CC fragments, lgs antisense DNA, lgs antisense RNA, lgs double-stranded RNA,
CC small peptides or chemical and natural compounds being capable of
CC interfering with Lgs function, synthesis and degradation. The disorders
CC are related to cell differentiation or cell proliferation. The compound
CC is administered to treat a cancerous condition by preventing progression
CC from a pre-neoplastic or non-malignant condition to a neoplastic or
CC malignant state. The cancerous condition is characterised by over-
CC stimulation of the Wnt pathway and is medulloblastoma or cancer of the
CC colon, breast, head and neck, brain, thyroid or skin. The therapeutic
CC compound may also be administered to a blood disease to promote tissue
CC regeneration and repair. This sequence represents the human Lgs/Bc19
CC polypeptide of the invention.
XX
SQ Sequence 1426 AA;

Query Match 100.0%; Score 183; DB 8; Length 1426;
Best Local Similarity 100.0%; Pred. No. 6e-16;
Matches 36; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 DGLSQEQLEHRERSLQTLRDIQRMLFPDEKEFTGAQ 36
||||||||||||||||||||||||||||||||||||
Db 349 DGLSQEQLEHRERSLQTLRDIQRMLFPDEKEFTGAQ 384

RESULT 6
ABB11808
ID ABB11808 standard; peptide; 1435 AA.
XX

AC ABB11808;
XX
DT 11-JAN-2002 (first entry)
XX
DE Human BCL9 homologue, SEQ ID NO:2178.
XX
KW Human; cytokine; cell proliferation; cell differentiation; growth factor;
KW haematopoiesis regulation; tissue growth; immunomodulator; activin;
KW inhibin; chemotaxis; chemokinesis; thrombolysis; oncogenesis;
KW proliferation; metastasis; cancer; tumour; haematopoietic disorder;
KW myeloid cell disorder; lymphoid cell disorder; asthma; arthritis;
KW chronic inflammatory condition; proliferative retinopathy;
KW atherosclerosis; coronary heart disease; arterial ischaemia;
KW bone disorder; osteoporosis; vascular growth disorder;
KW tissue regeneration; wound healing; infection; immune disorder;
KW cell culture; drug screening; gene therapy; antiinflammatory;
KW antiasthmatic; antiarthritic; haemostatic; antiarteriosclerotic;
KW cytostatic; osteopathic; vasotropic; cardiant; virucide; antibacterial;
KW antifungal; vulnerary; antiulcer.
XX
OS Homo sapiens.
XX
PN WO200157188-A2.
XX
PD 09-AUG-2001.
XX
PF 05-FEB-2001; 2001WO-US003800.
XX
PR 03-FEB-2000; 2000US-00496914.
PR 27-APR-2000; 2000US-00560875.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Drmanac RT;
XX
DR WPI; 2001-457740/49.
DR N-PSDB; ABA09052.
XX
PT Human proteins and DNA encoding sequences useful for preventing, treating
PT or ameliorating a medical condition in a mammalian subject e.g. arthritis
PT and cancer.
XX
PS Claim 20; Page 256-257; 1963pp; English.
XX
CC Sequences ABB10981-ABB12330 represent 1350 novel human polypeptides, and
CC sequences ABA08225-ABA09574 represent nucleic acids encoding them. The
CC invention also relates to vectors and recombinant host cells comprising a
CC nucleotide of the invention, methods of producing the novel polypeptides,
CC antibodies against the polypeptides, methods of detecting the nucleotides
CC or polypeptides in a sample, and methods of identifying compounds which
CC bind to polypeptides of the invention. Although novel, many of the
CC polypeptides of the invention have homology to known proteins, thereby
CC giving an insight into their probable biological activities, and hence
CC potential therapeutic applications. The polypeptides of the invention may
CC have various activities, including cytokine, cell proliferation or cell
CC differentiation activities; stem cell growth factor activity;
CC haematopoiesis regulatory activity; tissue growth activity;
CC immunomodulatory activity; activin- or inhibin-related activities;
CC chemotactic or chemokinetic activities; haemostatic, thrombotic or
CC thrombolytic activities; receptor or ligand activities; or may be
CC involved in oncogenesis, cancer cell proliferation or metastasis.
CC Depending on their biological activities, polypeptides and nucleotides of
CC the invention are useful for preventing, treating or ameliorating medical
CC conditions, e.g., by protein or gene therapy. Such conditions include
CC cancers, haematopoietic disorders (e.g.,myeloid or lymphoid cell
CC disorders), chronic inflammatory conditions (e.g., asthma or arthritis),
CC proliferative retinopathy, atherosclerosis, coronary heart disease,
CC arterial ischaemia, bone disorders (e.g., osteoporosis), and abnormal
CC vascular growth. Polypeptides involved with tissue regeneration and
CC repair (or nucleic acids encoding them) may be used to promote wound
CC healing (e.g., of burns, incisions and ulcers), while those with
CC immunomodulatory activities may be used in the treatment of viral,
CC bacterial and fungal infections in addition to immune disorders.
CC Polypeptides with growth factor activity may be used in cell cultures to
CC promote cell growth. For example, such polypeptides may be used to
CC manipulate stem cells in culture to give rise to neuroepithelial cells
CC that can be used to augment or replace cells damaged by illness,
CC autoimmune disease or accidental damage. The polypeptides and nucleotides
CC may also be used in the diagnosis of the above conditions, and in drug
CC screening techniques. The present sequence represents a novel human
CC polypeptide of the invention
XX
SQ Sequence 1435 AA;

Query Match 100.0%; Score 183; DB 4; Length 1435;
Best Local Similarity 100.0%; Pred. No. 6e-16;
Matches 36; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 DGLSQEQLEHRERSLQTLRDIQRMLFPDEKEFTGAQ 36
||||||||||||||||||||||||||||||||||||
Db 389 DGLSQEQLEHRERSLQTLRDIQRMLFPDEKEFTGAQ 424

RESULT 7
ADJ71893
ID ADJ71893 standard; peptide; 35 AA.
XX
AC ADJ71893;
XX
DT 20-MAY-2004 (first entry)
XX
DE Human Lgs/Bc19 peptide fragment #2.
XX
KW Human; legless; lgs; cell differentiation disorder;
KW cell proliferation disorder; cancer; Wnt pathway; medulloblastoma; colon;
KW breast; head; neck; brain; thyroid; skin; blood disease;
KW tissue regeneration; tissue repair; cytostatic; Lgs/Bc19.
XX
OS Homo sapiens.
XX
PN US2004038901-A1.
XX
PD 26-FEB-2004.
XX
PF 22-SEP-2003; 2003US-00664859.
XX
PR 28-JUL-2000; 2000US-0221502P.
PR 27-JUL-2001; 2001US-00915543.
XX
PA (UYZU-) UNIV ZURICH.
XX
PI Basler K, Brunner E, Froesch B, Kramps T, Peter O;
XX
DR WPI; 2004-203288/19.
XX
PT Novel polypeptide sharing one or more homologue amino acid domains with
PT Legless protein being functional homologue of Legless, useful for
PT diagnosing disorders of cell fate.
XX
PS Disclosure; SEQ ID NO 5; 62pp; English.
XX
CC The invention relates to a polypeptide sharing one or more homologous
CC amino acid domains with a Legless (Lgs) protein and is therefore a
CC functional homologue of Lgs. The invention also relates to a nucleotide
CC sequence encoding a protein present in invertebrate and/or vertebrate
CC organisms, the nucleotide sequence encoding a protein comprising a
CC positive function in a regulatory pathway and the use of the polypeptide
CC for the isolation of Lgs-binding proteins by carrying out an assay chosen
CC from an in vitro binding assay with such a peptide or a co-
CC immunoprecipitation from vertebrate or invertebrate cell lysates or a
CC mammalian or yeast two hybrid assay. The polypeptide and polynucleotide
CC are useful for treating disorders of cell fate, which involves
CC administering therapeutic compounds chosen from invertebrate and
CC vertebrate Lgs protein homologues or fragments, antibodies, antibody
CC fragments, lgs antisense DNA, lgs antisense RNA, lgs double-stranded RNA,
CC small peptides or chemical and natural compounds being capable of

```
CC   interfering with Lgs function, synthesis and degradation. The disorders
CC   are related to cell differentiation or cell proliferation. The compound
CC   is administered to treat a cancerous condition by preventing progression
CC   from a pre-neoplastic or non-malignant condition to a neoplastic or
CC   malignant state. The cancerous condition is characterised by over-
CC   stimulation of the Wnt pathway and is medulloblastoma or cancer of the
CC   colon, breast, head and neck, brain, thyroid or skin. The therapeutic
CC   compound may also be administered to a blood disease to promote tissue
CC   regeneration and repair. This sequence represents a human Lgs/Bc19
CC   peptide fragment of the invention.
XX
SQ   Sequence 35 AA;

  Query Match             97.3%;  Score 178;  DB 8;  Length 35;
  Best Local Similarity   100.0%;  Pred. No. 5e-17;
  Matches   35;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 DGLSQEQLEHRERSLQTLRDIQRMLFPDEKEFTGA 35
              |||||||||||||||||||||||||||||||||||
Db          1 DGLSQEQLEHRERSLQTLRDIQRMLFPDEKEFTGA 35


RESULT 8
AAU78461
ID   AAU78461 standard; protein; 320 AA.
XX
AC   AAU78461;
XX
DT   02-JUL-2002  (first entry)
XX
DE   Mouse beta-catenin nuclear localised protein #2.
XX
KW   Mouse; beta-catenin nuclear localised protein; cancer; gene therapy; EST;
KW   expressed sequence tag.
XX
OS   Mus musculus.
XX
PN   WO200224738-A1.
XX
PD   28-MAR-2002.
XX
PF   19-SEP-2001; 2001WO-JP008140.
XX
PR   22-SEP-2000; 2000JP-00287876.
XX
PA   (KYOW ) KYOWA HAKKO KOGYO KK.
XX
PI   Akiyama T,  Adachi S;
XX
DR   WPI; 2002-330014/36.
DR   N-PSDB; ABK47632.
XX
PT   New beta-catenin nuclear localized protein for diagnosis and treatment of
PT   diseases associated with nuclear localization of beta-catenin e.g.
PT   cancer.
XX
PS   Claim 2; Page 91-92; 113pp; Japanese.
XX
CC   The invention relates to a beta-catenin nuclear localised protein and DNA
CC   encoding the protein. The protein and encoding DNA are applicable in
CC   diagnosis and treatment of diseases associated with nuclear localisation
CC   of beta-catenin e.g. cancer, including gene therapy. The present sequence
CC   represents the amino acid sequence of mouse beta-catenin nuclear
CC   localised protein #2
XX
SQ   Sequence 320 AA;

  Query Match             59.6%;  Score 109;  DB 5;  Length 320;
  Best Local Similarity   84.0%;  Pred. No. 1.7e-06;
  Matches   21;  Conservative    4;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 DGLSQEQLEHRERSLQTLRDIQRML 25
              :|||:||||||||||||||||:|:|
Db        150 EGLSKEQLEHRERSLQTLRDIERLL 174


RESULT 9
AAU78460
ID   AAU78460 standard; protein; 1494 AA.
XX
AC   AAU78460;
XX
DT   02-JUL-2002  (first entry)
XX
DE   Mouse beta-catenin nuclear localised protein.
XX
KW   Mouse; beta-catenin nuclear localised protein; cancer; gene therapy; EST;
KW   expressed sequence tag.
XX
OS   Mus musculus.
XX
PN   WO200224738-A1.
XX
PD   28-MAR-2002.
XX
PF   19-SEP-2001; 2001WO-JP008140.
XX
PR   22-SEP-2000; 2000JP-00287876.
XX
PA   (KYOW ) KYOWA HAKKO KOGYO KK.
XX
PI   Akiyama T,  Adachi S;
XX
DR   WPI; 2002-330014/36.
DR   N-PSDB; ABK47631.
XX
PT   New beta-catenin nuclear localized protein for diagnosis and treatment of
PT   diseases associated with nuclear localization of beta-catenin e.g.
PT   cancer.
XX
PS   Claim 1; Page 81-88; 113pp; Japanese.
XX
CC   The invention relates to a beta-catenin nuclear localised protein and DNA
CC   encoding the protein. The protein and encoding DNA are applicable in
CC   diagnosis and treatment of diseases associated with nuclear localisation
CC   of beta-catenin e.g. cancer, including gene therapy. The present sequence
CC   represents the amino acid sequence of mouse beta-catenin nuclear
CC   localised protein
XX
SQ   Sequence 1494 AA;

  Query Match             59.6%;  Score 109;  DB 5;  Length 1494;
  Best Local Similarity   84.0%;  Pred. No. 9.3e-06;
  Matches   21;  Conservative    4;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 DGLSQEQLEHRERSLQTLRDIQRML 25
              :|||:||||||||||||||||:|:|
Db        394 EGLSKEQLEHRERSLQTLRDIERLL 418


RESULT 10
AAU78463
ID   AAU78463 standard; protein; 738 AA.
XX
AC   AAU78463;
XX
DT   02-JUL-2002  (first entry)
XX
DE   Human beta-catenin nuclear localised protein #2.
XX
KW   Mouse; beta-catenin nuclear localised protein; cancer; gene therapy; EST;
KW   expressed sequence tag.
XX
OS   Homo sapiens.
XX
PN   WO200224738-A1.
```

XX
PD 28-MAR-2002.
XX
PF 19-SEP-2001; 2001WO-JP008140.
XX
PR 22-SEP-2000; 2000JP-00287876.
XX
PA (KYOW) KYOWA HAKKO KOGYO KK.
XX
PI Akiyama T, Adachi S;
XX
DR WPI; 2002-330014/36.
DR N-PSDB; ABK47638.
XX
PT New beta-catenin nuclear localized protein for diagnosis and treatment of
PT diseases associated with nuclear localization of beta-catenin e.g.
PT cancer.
XX
PS Claim 8; Page 102-105; 113pp; Japanese.
XX
CC The invention relates to a beta-catenin nuclear localised protein and DNA
CC encoding the protein. The protein and encoding DNA are applicable in
CC diagnosis and treatment of diseases associated with nuclear localisation
CC of beta-catenin e.g. cancer, including gene therapy. The present sequence
CC represents the amino acid sequence of human beta-catenin nuclear
CC localised protein #2
XX
SQ Sequence 738 AA;

Query Match 58.5%; Score 107; DB 5; Length 738;
Best Local Similarity 87.5%; Pred. No. 8e-06;
Matches 21; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Qy 2 GLSQEQLEHRERSLQTLRDIQRML 25
||||:||||||||||||||:|:|
Db 1 GLSKEQLEHRERSLQTLRDIERLL 24

RESULT 11
AAB71230
ID AAB71230 standard; protein; 1115 AA.
XX
AC AAB71230;
XX
DT 18-NOV-2002 (first entry)
XX
DE Human legless homologue hlgs-1 partial protein.
XX
KW Legless; human; lgs; Wnt/Wingless signaling pathway; Wnt; Wg;
KW tissue proliferation; tumour; cytostatic; cellular disorder; colon;
KW blood disorder; cancer; breast; head and neck cancer; brain; thyroid;
KW medulloblastoma; skin cancer; tissue regeneration; tissue repair.
XX
OS Homo sapiens.
XX
PN US2002086986-A1.
XX
PD 04-JUL-2002.
XX
PF 27-JUL-2001; 2001US-00915543.
XX
PR 28-JUL-2000; 2000US-0221502P.
XX
PA (BASL/) BASLER K.
PA (BRUN/) BRUNNER E.
PA (FROE/) FROESCH B.
PA (KRAM/) KRAMPS T.
PA (PETE/) PETER O.
XX
PI Basler K, Brunner E, Froesch B, Kramps T, Peter O;
XX
DR WPI; 2002-635689/68.
DR N-PSDB; AAF88468.
XX
PT Novel polypeptide useful in therapeutic method for treating disorders of
PT cell fate such as cell differentiation or cell proliferation.
XX
PS Example II; Fig 10B; 41pp; English.
XX
CC This invention describes a novel polypeptide sharing one or more
CC homologous amino acid domains with the legless (lgs) protein, a
CC downstream component of the Wnt/Wingless (Wnt/Wg) signaling pathway
CC involved in the formation and maintenance of spatial arrangements and
CC proliferation of tissues during development, and in the formation and
CC growth of many human tumours. The products of the invention have
CC cytostatic activity and can be used to treat cellular disorders, blood
CC disorders and cancers caused by over-stimulation of the Wnt pathway,
CC where the cancerous condition is colon, breast, head and neck, brain,
CC thyroid, medulloblastoma or skin cancer. The product could also be used
CC to promote tissue regeneration and repair. This sequence represents the
CC human legless (lgs) protein homologue hlgs-1 described in the disclosure
CC of the invention
XX
SQ Sequence 1115 AA;

Query Match 58.5%; Score 107; DB 5; Length 1115;
Best Local Similarity 87.5%; Pred. No. 1.3e-05;
Matches 21; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Qy 2 GLSQEQLEHRERSLQTLRDIQRML 25
||||:||||||||||||||:|:|
Db 76 GLSKEQLEHRERSLQTLRDIERLL 99

RESULT 12
ABW01535
ID ABW01535 standard; protein; 1115 AA.
XX
AC ABW01535;
XX
DT 15-JAN-2004 (first entry)
XX
DE Human lgs-1 protein.
XX
KW Legless protein; lgs; cell fate disorder; blood disease; gene therapy;
KW cancer; tissue regeneration; tissue repair; cytostatic.
XX
OS Homo sapiens.
XX
PN US2003114413-A1.
XX
PD 19-JUN-2003.
XX
PF 19-DEC-2002; 2002US-00322579.
XX
PR 28-JUL-2000; 2000US-0221502P.
PR 27-JUL-2001; 2001US-00915543.
XX
PA (UYZU-) UNIV ZURICH.
XX
PI Basler K, Brunner E, Froesch B, Kramps T, Peter O;
XX
DR WPI; 2003-829432/77.
DR N-PSDB; AAD62643.
XX
PT Novel Lgs polypeptide useful for isolation of Lgs-binding proteins,
PT diagnosing disorders of cell fate, treating diseases such as cancer.
XX
PS Claim 7; Fig 10B; 0pp; English.
XX
CC The invention relates to novel legless (lgs) proteins and polynucleotides
CC encoding such proteins. Lgs sequences are useful for the treatment of
CC disorders of cell fate such as differentiation or proliferation. The
CC invention is used to treat blood disease or a cancerous condition
CC characterised by over-stimulation of the Wnt pathway such as colon,
CC breast, head and neck, brain, thyroid, medulloblastoma or skin cancer and

CC is administered to prevent progression from a pre-neoplastic or non-
CC malignant condition to a neoplastic or malignant state. It is
CC administered to promote tissue regeneration and repair. The invention is
CC also useful in the therapy of diseases cost by an over-activation of Wg
CC pathway. It is useful for reducing lgs gene expression in an invertebrate
CC or vertebrate organism or an invertebrate or vertebrate cell line. The
CC invention is also useful in gene therapy. The present sequence is human
CC lgs-1 protein used in the invention
XX
SQ Sequence 1115 AA;

Query Match 58.5%; Score 107; DB 7; Length 1115;
Best Local Similarity 87.5%; Pred. No. 1.3e-05;
Matches 21; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Qy 2 GLSQEQLEHRERSLQTLRDIQRML 25
||||:||||||||||||||:|:|
Db 76 GLSKEQLEHRERSLQTLRDIERLL 99

RESULT 13
ADJ71905
ID ADJ71905 standard; protein; 1115 AA.
XX
AC ADJ71905;
XX
DT 20-MAY-2004 (first entry)
XX
DE Human Lgs/Bc19 partial polypeptide.
XX
KW Human; legless; lgs; cell differentiation disorder;
KW cell proliferation disorder; cancer; Wnt pathway; medulloblastoma; colon;
KW breast; head; neck; brain; thyroid; skin; blood disease;
KW tissue regeneration; tissue repair; cytostatic; Lgs/Bc19.
XX
OS Homo sapiens.
XX
PN US2004038901-A1.
XX
PD 26-FEB-2004.
XX
PF 22-SEP-2003; 2003US-00664859.
XX
PR 28-JUL-2000; 2000US-0221502P.
PR 27-JUL-2001; 2001US-00915543.
XX
PA (UYZU-) UNIV ZURICH.
XX
PI Basler K, Brunner E, Froesch B, Kramps T, Peter O;
XX
DR WPI; 2004-203288/19.
DR N-PSDB; ADJ71904.
XX
PT Novel polypeptide sharing one or more homologue amino acid domains with
PT Legless protein being functional homologue of Legless, useful for
PT diagnosing disorders of cell fate.
XX
PS Example 2; SEQ ID NO 17; 62pp; English.
XX
CC The invention relates to a polypeptide sharing one or more homologous
CC amino acid domains with a Legless (Lgs) protein and is therefore a
CC functional homologue of Lgs. The invention also relates to a nucleotide
CC sequence encoding a protein present in invertebrate and/or vertebrate
CC organisms, the nucleotide sequence encoding a protein comprising a
CC positive function in a regulatory pathway and the use of the polypeptide
CC for the isolation of Lgs-binding proteins by carrying out an assay chosen
CC from an in vitro binding assay with such a peptide or a co-
CC immunoprecipitation from vertebrate or invertebrate cell lysates or a
CC mammalian or yeast two hybrid assay. The polypeptide and polynucleotide
CC are useful for treating disorders of cell fate, which involves
CC administering therapeutic compounds chosen from invertebrate and
CC vertebrate Lgs protein homologues or fragments, antibodies, antibody
CC fragments, lgs antisense DNA, lgs antisense RNA, lgs double-stranded RNA,
CC small peptides or chemical and natural compounds being capable of
CC interfering with Lgs function, synthesis and degradation. The disorders
CC are related to cell differentiation or cell proliferation. The compound
CC is administered to treat a cancerous condition by preventing progression
CC from a pre-neoplastic or non-malignant condition to a neoplastic or
CC malignant state. The cancerous condition is characterised by over-
CC stimulation of the Wnt pathway and is medulloblastoma or cancer of the
CC colon, breast, head and neck, brain, thyroid or skin. The therapeutic
CC compound may also be administered to a blood disease to promote tissue
CC regeneration and repair. This sequence represents a human Lgs/Bc19
CC partial polypeptide of the invention.
XX
SQ Sequence 1115 AA;

Query Match 58.5%; Score 107; DB 8; Length 1115;
Best Local Similarity 87.5%; Pred. No. 1.3e-05;
Matches 21; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Qy 2 GLSQEQLEHRERSLQTLRDIQRML 25
||||:||||||||||||||:|:|
Db 76 GLSKEQLEHRERSLQTLRDIERLL 99

RESULT 14
ABW01529
ID ABW01529 standard; peptide; 35 AA.
XX
AC ABW01529;
XX
DT 15-JAN-2004 (first entry)
XX
DE Drosophila species legless (lgs) peptide #2.
XX
KW Legless protein; lgs; cell fate disorder; blood disease; gene therapy;
KW cancer; tissue regeneration; tissue repair; cytostatic.
XX
OS Drosophila sp.
XX
PN US2003114413-A1.
XX
PD 19-JUN-2003.
XX
PF 19-DEC-2002; 2002US-00322579.
XX
PR 28-JUL-2000; 2000US-0221502P.
PR 27-JUL-2001; 2001US-00915543.
XX
PA (UYZU-) UNIV ZURICH.
XX
PI Basler K, Brunner E, Froesch B, Kramps T, Peter O;
XX
DR WPI; 2003-829432/77.
XX
PT Novel Lgs polypeptide useful for isolation of Lgs-binding proteins,
PT diagnosing disorders of cell fate, treating diseases such as cancer.
XX
PS Claim 28; Fig 7B; 0pp; English.
XX
CC The invention relates to novel legless (lgs) proteins and polynucleotides
CC encoding such proteins. Lgs sequences are useful for the treatment of
CC disorders of cell fate such as differentiation or proliferation. The
CC invention is used to treat blood disease or a cancerous condition
CC characterised by over-stimulation of the Wnt pathway such as colon,
CC breast, head and neck, brain, thyroid, medulloblastoma or skin cancer and
CC is administered to prevent progression from a pre-neoplastic or non-
CC malignant condition to a neoplastic or malignant state. It is
CC administered to promote tissue regeneration and repair. The invention is
CC also useful in the therapy of diseases cost by an over-activation of Wg
CC pathway. It is useful for reducing lgs gene expression in an invertebrate
CC or vertebrate organism or an invertebrate or vertebrate cell line. The
CC invention is also useful in gene therapy. The present sequence is
CC Drosophila species legless (lgs) peptide
XX

```
SQ   Sequence 35 AA;

  Query Match             36.1%;  Score 66;  DB 7;  Length 35;
  Best Local Similarity   31.4%;  Pred. No. 0.12;
  Matches   11;  Conservative   10;  Mismatches   14;  Indels    0;  Gaps    0;

Qy          1 DGLSQEQLEHRERSLQTLRDIQRMLFPDEKEFTGA 35
              : |: :| :|||  |  :: : : |||: :   ||
Db          1 ENLTPQQRQHREEQLAKIKKMNQFLFPENENSVGA 35


RESULT 15
ADJ71892
ID   ADJ71892 standard; peptide; 35 AA.
XX
AC   ADJ71892;
XX
DT   20-MAY-2004  (first entry)
XX
DE   Fruit fly legless (lgs) peptide fragment #2.
XX
KW   Fruit fly; legless; lgs; cell differentiation disorder;
KW   cell proliferation disorder; cancer; Wnt pathway; medulloblastoma; colon;
KW   breast; head; neck; brain; thyroid; skin; blood disease;
KW   tissue regeneration; tissue repair; cytostatic.
XX
OS   Drosophila melanogaster.
XX
PN   US2004038901-A1.
XX
PD   26-FEB-2004.
XX
PF   22-SEP-2003; 2003US-00664859.
XX
PR   28-JUL-2000; 2000US-0221502P.
PR   27-JUL-2001; 2001US-00915543.
XX
PA   (UYZU-) UNIV ZURICH.
XX
PI   Basler K,  Brunner E,  Froesch B,  Kramps T,  Peter O;
XX
DR   WPI; 2004-203288/19.
XX
PT   Novel polypeptide sharing one or more homologue amino acid domains with
PT   Legless protein being functional homologue of Legless, useful for
PT   diagnosing disorders of cell fate.
XX
PS   Disclosure; SEQ ID NO 4; 62pp; English.
XX
CC   The invention relates to a polypeptide sharing one or more homologous
CC   amino acid domains with a Legless (Lgs) protein and is therefore a
CC   functional homologue of Lgs. The invention also relates to a nucleotide
CC   sequence encoding a protein present in invertebrate and/or vertebrate
CC   organisms, the nucleotide sequence encoding a protein comprising a
CC   positive function in a regulatory pathway and the use of the polypeptide
CC   for the isolation of Lgs-binding proteins by carrying out an assay chosen
CC   from an in vitro binding assay with such a peptide or a co-
CC   immunoprecipitation from vertebrate or invertebrate cell lysates or a
CC   mammalian or yeast two hybrid assay. The polypeptide and polynucleotide
CC   are useful for treating disorders of cell fate, which involves
CC   administering therapeutic compounds chosen from invertebrate and
CC   vertebrate Lgs protein homologues or fragments, antibodies, antibody
CC   fragments, lgs antisense DNA, lgs antisense RNA, lgs double-stranded RNA,
CC   small peptides or chemical and natural compounds being capable of
CC   interfering with Lgs function, synthesis and degradation. The disorders
CC   are related to cell differentiation or cell proliferation. The compound
CC   is administered to treat a cancerous condition by preventing progression
CC   from a pre-neoplastic or non-malignant condition to a neoplastic or
CC   malignant state. The cancerous condition is characterised by over-
CC   stimulation of the Wnt pathway and is medulloblastoma or cancer of the
CC   colon, breast, head and neck, brain, thyroid or skin. The therapeutic
CC   compound may also be administered to a blood disease to promote tissue
CC   regeneration and repair. This sequence represents a Drosophila Lgs
CC   peptide fragment of the invention.
XX
SQ   Sequence 35 AA;

  Query Match             36.1%;  Score 66;  DB 8;  Length 35;
  Best Local Similarity   31.4%;  Pred. No. 0.12;
  Matches   11;  Conservative   10;  Mismatches   14;  Indels    0;  Gaps    0;

Qy          1 DGLSQEQLEHRERSLQTLRDIQRMLFPDEKEFTGA 35
              : |: :| :|||  |  :: : : |||: :   ||
Db          1 ENLTPQQRQHREEQLAKIKKMNQFLFPENENSVGA 35


RESULT 16
ABB58779
ID   ABB58779 standard; protein; 1429 AA.
XX
AC   ABB58779;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Drosophila melanogaster polypeptide SEQ ID NO 3129.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US009231.
XX
PR   23-MAR-2000; 2000US-0191637P.
PR   11-JUL-2000; 2000US-00614150.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Venter JC,  Adams M,  Li PWD,  Myers EW;
XX
DR   WPI; 2001-656860/75.
DR   N-PSDB; ABL02882.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signaling and cell-cell
PT   interactions.
XX
PS   Disclosure; SEQ ID NO 3129; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-
CC   ABB72072). The sequence data for this patent did not form part of the
CC   printed specification, but was obtained in electronic format directly
CC   from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 1429 AA;

  Query Match             36.1%;  Score 66;  DB 4;  Length 1429;
  Best Local Similarity   31.4%;  Pred. No. 7.1;
  Matches   11;  Conservative   10;  Mismatches   14;  Indels    0;  Gaps    0;

Qy          1 DGLSQEQLEHRERSLQTLRDIQRMLFPDEKEFTGA 35
              : |: :| :|||  |  :: : : |||: :   ||
Db        520 ENLTPQQRQHREEQLAKIKKMNQFLFPENENSVGA 554


RESULT 17
```

```
AAB71228
ID   AAB71228 standard; protein; 1464 AA.
XX
AC   AAB71228;
XX
DT   18-NOV-2002  (first entry)
XX
DE   D. melanogaster lgs protein.
XX
KW   Legless; fruitfly; lgs; Wnt/Wingless signaling pathway; Wnt; Wg;
KW   tissue proliferation; tumour; cytostatic; cellular disorder; colon;
KW   blood disorder; cancer; breast; head and neck cancer; brain; thyroid;
KW   medulloblastoma; skin cancer; tissue regeneration; tissue repair.
XX
OS   Drosophila melanogaster.
XX
PN   US2002086986-A1.
XX
PD   04-JUL-2002.
XX
PF   27-JUL-2001; 2001US-00915543.
XX
PR   28-JUL-2000; 2000US-0221502P.
XX
PA   (BASL/) BASLER K.
PA   (BRUN/) BRUNNER E.
PA   (FROE/) FROESCH B.
PA   (KRAM/) KRAMPS T.
PA   (PETE/) PETER O.
XX
PI   Basler K,  Brunner E,  Froesch B,  Kramps T,  Peter O;
XX
DR   WPI; 2002-635689/68.
DR   N-PSDB; AAF88466.
XX
PT   Novel polypeptide useful in therapeutic method for treating disorders of
PT   cell fate such as cell differentiation or cell proliferation.
XX
PS   Example II; Fig 2; 41pp; English.
XX
CC   This invention describes a novel polypeptide sharing one or more
CC   homologous amino acid domains with the legless (lgs) protein, a
CC   downstream component of the Wnt/Wingless (Wnt/Wg) signaling pathway
CC   involved in the formation and maintenance of spatial arrangements and
CC   proliferation of tissues during development, and in the formation and
CC   growth of many human tumours. The products of the invention have
CC   cytostatic activity and can be used to treat cellular disorders, blood
CC   disorders and cancers caused by over-stimulation of the Wnt pathway,
CC   where the cancerous condition is colon, breast, head and neck, brain,
CC   thyroid, medulloblastoma or skin cancer. The product could also be used
CC   to promote tissue regeneration and repair. This sequence represents the
CC   Drosophila melanogaster (fruitfly) legless (lgs) protein described in the
CC   disclosure of the invention
XX
SQ   Sequence 1464 AA;

  Query Match             36.1%;  Score 66;  DB 5;  Length 1464;
  Best Local Similarity   31.4%;  Pred. No. 7.3;
  Matches   11;  Conservative   10;  Mismatches   14;  Indels    0;  Gaps    0;

Qy          1 DGLSQEQLEHRERSLQTLRDIQRMLFPDEKEFTGA 35
              : |: :| :|||  |  :: : : |||: :   ||
Db        515 ENLTPQQRQHREEQLAKIKKMNQFLFPENENSVGA 549


RESULT 18
ABW01527
ID   ABW01527 standard; protein; 1464 AA.
XX
AC   ABW01527;
XX
DT   15-JAN-2004  (first entry)
XX
DE   Drosophila species legless (lgs) protein.
XX
KW   Legless protein; lgs; cell fate disorder; blood disease; gene therapy;
KW   cancer; tissue regeneration; tissue repair; cytostatic.
XX
OS   Drosophila sp.
XX
PN   US2003114413-A1.
XX
PD   19-JUN-2003.
XX
PF   19-DEC-2002; 2002US-00322579.
XX
PR   28-JUL-2000; 2000US-0221502P.
PR   27-JUL-2001; 2001US-00915543.
XX
PA   (UYZU-) UNIV ZURICH.
XX
PI   Basler K,  Brunner E,  Froesch B,  Kramps T,  Peter O;
XX
DR   WPI; 2003-829432/77.
DR   N-PSDB; AAD62641.
XX
PT   Novel Lgs polypeptide useful for isolation of Lgs-binding proteins,
PT   diagnosing disorders of cell fate, treating diseases such as cancer.
XX
PS   Claim 5; Fig 2; 0pp; English.
XX
CC   The invention relates to novel legless (lgs) proteins and polynucleotides
CC   encoding such proteins. Lgs sequences are useful for the treatment of
CC   disorders of cell fate such as differentiation or proliferation. The
CC   invention is used to treat blood disease or a cancerous condition
CC   characterised by over-stimulation of the Wnt pathway such as colon,
CC   breast, head and neck, brain, thyroid, medulloblastoma or skin cancer and
CC   is administered to prevent progression from a pre-neoplastic or non-
CC   malignant condition to a neoplastic or malignant state. It is
CC   administered to promote tissue regeneration and repair. The invention is
CC   also useful in the therapy of diseases cost by an over-activation of Wg
CC   pathway. It is useful for reducing lgs gene expression in an invertebrate
CC   or vertebrate organism or an invertebrate or vertebrate cell line. The
CC   invention is also useful in gene therapy. The present sequence is
CC   Drosophila species legless (lgs) protein
XX
SQ   Sequence 1464 AA;

  Query Match             36.1%;  Score 66;  DB 7;  Length 1464;
  Best Local Similarity   31.4%;  Pred. No. 7.3;
  Matches   11;  Conservative   10;  Mismatches   14;  Indels    0;  Gaps    0;

Qy          1 DGLSQEQLEHRERSLQTLRDIQRMLFPDEKEFTGA 35
              : |: :| :|||  |  :: : : |||: :   ||
Db        515 ENLTPQQRQHREEQLAKIKKMNQFLFPENENSVGA 549


RESULT 19
ADJ71911
ID   ADJ71911 standard; protein; 1464 AA.
XX
AC   ADJ71911;
XX
DT   20-MAY-2004  (first entry)
XX
DE   Fruit fly legless (lgs) polypeptide.
XX
KW   Fruit fly; legless; lgs; cell differentiation disorder;
KW   cell proliferation disorder; cancer; Wnt pathway; medulloblastoma; colon;
KW   breast; head; neck; brain; thyroid; skin; blood disease;
KW   tissue regeneration; tissue repair; cytostatic.
XX
OS   Drosophila melanogaster.
XX
PN   US2004038901-A1.
XX
```

PD 26-FEB-2004.
XX
PF 22-SEP-2003; 2003US-00664859.
XX
PR 28-JUL-2000; 2000US-0221502P.
PR 27-JUL-2001; 2001US-00915543.
XX
PA (UYZU-) UNIV ZURICH.
XX
PI Basler K, Brunner E, Froesch B, Kramps T, Peter O;
XX
DR WPI; 2004-203288/19.
DR N-PSDB; ADJ71889.
XX
PT Novel polypeptide sharing one or more homologue amino acid domains with
PT Legless protein being functional homologue of Legless, useful for
PT diagnosing disorders of cell fate.
XX
PS Example 2; SEQ ID NO 1; 62pp; English.
XX
CC The invention relates to a polypeptide sharing one or more homologous
CC amino acid domains with a Legless (Lgs) protein and is therefore a
CC functional homologue of Lgs. The invention also relates to a nucleotide
CC sequence encoding a protein present in invertebrate and/or vertebrate
CC organisms, the nucleotide sequence encoding a protein comprising a
CC positive function in a regulatory pathway and the use of the polypeptide
CC for the isolation of Lgs-binding proteins by carrying out an assay chosen
CC from an in vitro binding assay with such a peptide or a co-
CC immunoprecipitation from vertebrate or invertebrate cell lysates or a
CC mammalian or yeast two hybrid assay. The polypeptide and polynucleotide
CC are useful for treating disorders of cell fate, which involves
CC administering therapeutic compounds chosen from invertebrate and
CC vertebrate Lgs protein homologues or fragments, antibodies, antibody
CC fragments, lgs antisense DNA, lgs antisense RNA, lgs double-stranded RNA,
CC small peptides or chemical and natural compounds being capable of
CC interfering with Lgs function, synthesis and degradation. The disorders
CC are related to cell differentiation or cell proliferation. The compound
CC is administered to treat a cancerous condition by preventing progression
CC from a pre-neoplastic or non-malignant condition to a neoplastic or
CC malignant state. The cancerous condition is characterised by over-
CC stimulation of the Wnt pathway and is medulloblastoma or cancer of the
CC colon, breast, head and neck, brain, thyroid or skin. The therapeutic
CC compound may also be administered to a blood disease to promote tissue
CC regeneration and repair. This sequence represents the Drosophila Lgs
CC polypeptide of the invention.
XX
SQ Sequence 1464 AA;

Query Match 36.1%; Score 66; DB 8; Length 1464;
Best Local Similarity 31.4%; Pred. No. 7.3;
Matches 11; Conservative 10; Mismatches 14; Indels 0; Gaps 0;

Qy 1 DGLSQEQLEHRERSLQTLRDIQRMLFPDEKEFTGA 35
: |: :| :||| | :: : : |||: : ||
Db 515 ENLTPQQRQHREEQLAKIKKMNQFLFPENENSVGA 549

RESULT 20
ABP98879
ID ABP98879 standard; protein; 1014 AA.
XX
AC ABP98879;
XX
DT 24-JUL-2003 (first entry)
XX
DE Human molecule for disease detection and treatment MDDT-7.
XX
KW Cytostatic; antiarteriosclerotic; anti-HIV; antiallergic; nephrotropic;
KW antithyroid; cerebroprotective; antiparkinsonian; anticonvulsant; MDDT;
KW nootropic; neuroprotective; antidiabetic; gene therapy; atherosclerosis;
KW molecule for disease detection and treatment; cancer; AIDS; allergy;
KW diabetes; glomerulonephritis; autoimmune thyroiditis; Cushing's syndrome;
KW stroke; Parkinson's disease; epilepsy.
XX
OS Homo sapiens.
XX
PN WO2003031595-A2.
XX
PD 17-APR-2003.
XX
PF 10-OCT-2002; 2002WO-US032852.
XX
PR 12-OCT-2001; 2001US-0328944P.
PR 26-OCT-2001; 2001US-0345384P.
PR 02-NOV-2001; 2001US-0343880P.
PR 09-NOV-2001; 2001US-0345143P.
PR 16-NOV-2001; 2001US-0332430P.
XX
PA (INCY-) INCYTE GENOMICS INC.
XX
PI Tang YT, Forsythe IJ, Emerling BM, Hafalia AJA, Yue H, Xu Y;
PI Gietzen KJ, Chawla NK, Baughn MR, Marquis JP, Becha SD, Kable AE;
PI Lal PG, Richardson TW, Lee SY, Lee EA, Tran B, Warren BA, Lu DAM;
PI Gururajan R, Sprague WW, Blake JJ, Thangavelu K, Swarnakar A;
PI Gorvad AE, Griffin JA, Lindquist EA, Elliott VS, Ison CH;
PI Ramkumar J;
XX
DR WPI; 2003-421277/39.
DR N-PSDB; ACC44394.
XX
PT Isolated peptide molecules for disease detection and treatment, useful
PT for diagnosing, treating or preventing disorders, e.g. cancer, AIDS,
PT atherosclerosis, diabetes or stroke.
XX
PS Claim 1; Page 151-153; 234pp; English.
XX
CC The invention relates to the isolation of a number of "molecules for
CC disease detection and treatment" (MDDT) and genes encoding them. The
CC invention also includes molecule which are at least 90% identical to the
CC protein and nucleotide sequences. This sequence represents a protein of
CC the invention. Disorders associated with aberrant expression of MDDT, are
CC cell proliferative disorders (e.g. cancer or atherosclerosis),
CC autoimmune/inflammatory disorders (e.g. AIDS, allergies, diabetes,
CC glomerulonephritis or autoimmune thyroiditis), developmental disorders
CC (e.g. Cushing's syndrome) or neurological disorders (e.g. stroke,
CC Parkinson's disease or epilepsy)
XX
SQ Sequence 1014 AA;

Query Match 33.6%; Score 61.5; DB 6; Length 1014;
Best Local Similarity 44.8%; Pred. No. 20;
Matches 13; Conservative 10; Mismatches 5; Indels 1; Gaps 1;

Qy 8 LEHRERSLQTLRDIQRMLFPDEKEFTGAQ 36
::|::|:: |:||: :| | |:|||||
Db 51 VKHKDRTM-NLQDIRYILKNDLKDFTGAQ 78

RESULT 21
ABJ25853
ID ABJ25853 standard; protein; 603 AA.
XX
AC ABJ25853;
XX
DT 16-APR-2003 (first entry)
XX
DE Aspergillus fumigatus essential gene protein #511.
XX
KW Fungicide; cytostatic; essential gene; Aspergillus fumigatus; infection;
KW cancer; contamination; biofilm; antibody; immune response.
XX
OS Aspergillus fumigatus.
XX
PN WO200286090-A2.
XX
PD 31-OCT-2002.

```
XX
PF   23-APR-2002; 2002WO-US013142.
XX
PR   23-APR-2001; 2001US-0285697P.
PR   27-APR-2001; 2001US-0287066P.
PR   05-JUN-2001; 2001US-0295890P.
PR   09-JUL-2001; 2001US-0303899P.
PR   31-AUG-2001; 2001US-0316362P.
XX
PA   (ELIT-) ELITRA PHARM INC.
XX
PI   Jiang B,  Tishkoff D,  Zamudio C,  Eroshkin AM,  Hu W,  Lemieux SM;
XX
DR   WPI; 2003-093124/08.
XX
PT   New purified or isolated nucleic acids of essential genes of Aspergillus
PT   fumigatus, useful for treating or preventing infections by A. fumigatus,
PT   or for treating a non-infectious disease in a subject e.g. cancer.
XX
PS   Disclosure; Page; 175pp; English.
XX
CC   The invention relates to novel purified or isolated nucleic acids of
CC   essential genes of Aspergillus fumigatus. The isolated nucleic acids of
CC   the invention are used to treat or prevent infections by a pathogenic
CC   organism such as A. fumigatus, to treat a non-infectious disease in a
CC   subject (e.g. cancer), to prevent or contain contamination of an object
CC   by A. fumigatus, or to prevent or inhibit formation on a surface of a
CC   biofilm comprising A. fumigatus. The polynucleotides are useful for
CC   expressing recombinant protein for characterisation, screening or
CC   therapeutic use, as markers for host tissues in which the pathogenic
CC   organisms invade or reside, for comparing with the DNA sequence of A.
CC   fumigatus to identify duplicated genes or paralogues having the same or
CC   similar biochemical activity and/or function, for comparing with DNA
CC   sequences of other related or distant pathogenic organisms to identify
CC   potential orthologous essential or virulence genes, for selecting and
CC   making oligomers for attachment to a nucleic acid array for examination
CC   of expression patterns, for raising anti-protein antibodies, as an
CC   antigen to raise anti-DNA antibodies or to elicit another immune
CC   response, and for identifying polynucleotides encoding the other protein
CC   with which binding occurs or to identify inhibitors of the binding
CC   interaction. The polypeptides may be used to raise antibodies or to
CC   elicit immune response, as a reagent in assays designed to quantitatively
CC   determine levels of the protein in biological fluids, as a marker for
CC   host tissues in which pathogenic organism invade or reside, and to
CC   isolate correlative receptors or ligands in the case or virulence
CC   factors. This sequence represents a protein of one of the essential genes
CC   of Aspergillus fumigatus of the invention
XX
SQ   Sequence 603 AA;

  Query Match             31.4%;  Score 57.5;  DB 6;  Length 603;
  Best Local Similarity   31.4%;  Pred. No. 41;
  Matches   11;  Conservative   11;  Mismatches   10;  Indels    3;  Gaps    1;

Qy          1 DGLSQEQLEHRERSLQTLRDIQRMLFPDEKEFTGA 35
              ||:  |::  ::   :  : ::|||| |:: | ||
Db         45 DGVETEKIREKD---EVEKKLERMLFGDDEGFVGA 76


RESULT 22
ABJ26453
ID   ABJ26453 standard; protein; 618 AA.
XX
AC   ABJ26453;
XX
DT   16-APR-2003  (first entry)
XX
DE   Aspergillus fumigatus essential gene protein #1111.
XX
KW   Fungicide; cytostatic; essential gene; Aspergillus fumigatus; infection;
KW   cancer; contamination; biofilm; antibody; immune response.
XX
OS   Aspergillus fumigatus.
XX
PN   WO200286090-A2.
XX
PD   31-OCT-2002.
XX
PF   23-APR-2002; 2002WO-US013142.
XX
PR   23-APR-2001; 2001US-0285697P.
PR   27-APR-2001; 2001US-0287066P.
PR   05-JUN-2001; 2001US-0295890P.
PR   09-JUL-2001; 2001US-0303899P.
PR   31-AUG-2001; 2001US-0316362P.
XX
PA   (ELIT-) ELITRA PHARM INC.
XX
PI   Jiang B,  Tishkoff D,  Zamudio C,  Eroshkin AM,  Hu W,  Lemieux SM;
XX
DR   WPI; 2003-093124/08.
XX
PT   New purified or isolated nucleic acids of essential genes of Aspergillus
PT   fumigatus, useful for treating or preventing infections by A. fumigatus,
PT   or for treating a non-infectious disease in a subject e.g. cancer.
XX
PS   Disclosure; Page; 175pp; English.
XX
CC   The invention relates to novel purified or isolated nucleic acids of
CC   essential genes of Aspergillus fumigatus. The isolated nucleic acids of
CC   the invention are used to treat or prevent infections by a pathogenic
CC   organism such as A. fumigatus, to treat a non-infectious disease in a
CC   subject (e.g. cancer), to prevent or contain contamination of an object
CC   by A. fumigatus, or to prevent or inhibit formation on a surface of a
CC   biofilm comprising A. fumigatus. The polynucleotides are useful for
CC   expressing recombinant protein for characterisation, screening or
CC   therapeutic use, as markers for host tissues in which the pathogenic
CC   organisms invade or reside, for comparing with the DNA sequence of A.
CC   fumigatus to identify duplicated genes or paralogues having the same or
CC   similar biochemical activity and/or function, for comparing with DNA
CC   sequences of other related or distant pathogenic organisms to identify
CC   potential orthologous essential or virulence genes, for selecting and
CC   making oligomers for attachment to a nucleic acid array for examination
CC   of expression patterns, for raising anti-protein antibodies, as an
CC   antigen to raise anti-DNA antibodies or to elicit another immune
CC   response, and for identifying polynucleotides encoding the other protein
CC   with which binding occurs or to identify inhibitors of the binding
CC   interaction. The polypeptides may be used to raise antibodies or to
CC   elicit immune response, as a reagent in assays designed to quantitatively
CC   determine levels of the protein in biological fluids, as a marker for
CC   host tissues in which pathogenic organism invade or reside, and to
CC   isolate correlative receptors or ligands in the case or virulence
CC   factors. This sequence represents a protein of one of the essential genes
CC   of Aspergillus fumigatus of the invention
XX
SQ   Sequence 618 AA;

  Query Match             31.4%;  Score 57.5;  DB 6;  Length 618;
  Best Local Similarity   31.4%;  Pred. No. 42;
  Matches   11;  Conservative   11;  Mismatches   10;  Indels    3;  Gaps    1;

Qy          1 DGLSQEQLEHRERSLQTLRDIQRMLFPDEKEFTGA 35
              ||:  |::  ::   :  : ::|||| |:: | ||
Db         45 DGVETEKIREKD---EVEKKLERMLFGDDEGFVGA 76


RESULT 23
ABU44941
ID   ABU44941 standard; protein; 411 AA.
XX
AC   ABU44941;
XX
DT   19-JUN-2003  (first entry)
XX
DE   Protein encoded by Prokaryotic essential gene #30468.
XX
```

```
KW   Antisense; prokaryotic essential gene; cell proliferation; drug design.
XX
OS   Salmonella paratyphi.
XX
PN   WO200277183-A2.
XX
PD   03-OCT-2002.
XX
PF   21-MAR-2002; 2002WO-US009107.
XX
PR   21-MAR-2001; 2001US-00815242.
PR   06-SEP-2001; 2001US-00948993.
PR   25-OCT-2001; 2001US-0342923P.
PR   08-FEB-2002; 2002US-00072851.
PR   06-MAR-2002; 2002US-0362699P.
XX
PA   (ELIT-) ELITRA PHARM INC.
XX
PI   Wang L,  Zamudio C,  Malone C,  Haselbeck R,  Ohlsen KL,  Zyskind JW;
PI   Wall D,  Trawick JD,  Carr GJ,  Yamamoto R,  Forsyth RA,  Xu HH;
XX
DR   WPI; 2003-029926/02.
DR   N-PSDB; ACA48811.
XX
PT   New antisense nucleic acids, useful for identifying proteins or screening
PT   for homologous nucleic acids required for cellular proliferation to
PT   isolate candidate molecules for rational drug discovery programs.
XX
PS   Claim 25; SEQ ID NO 72865; 1766pp; English.
XX
CC   The invention relates to an isolated nucleic acid comprising any one of
CC   the 6213 antisense sequences given in the specification where expression
CC   of the nucleic acid inhibits proliferation of a cell. Also included are:
CC   (1) a vector comprising a promoter operably linked to the nucleic acid
CC   encoding a polypeptide whose expression is inhibited by the antisense
CC   nucleic acid; (2) a host cell containing the vector; (3) an isolated
CC   polypeptide or its fragment whose expression is inhibited by the
CC   antisense nucleic acid; (4) an antibody capable of specifically binding
CC   the polypeptide;(5) producing the polypeptide; (6) inhibiting cellular
CC   proliferation or the activity of a gene in an operon required for
CC   proliferation; (7) identifying a compound that influences the activity of
CC   the gene product or that has an activity against a biological pathway
CC   required for proliferation, or that inhibits cellular proliferation; (8)
CC   identifying a gene required for cellular proliferation or the biological
CC   pathway in which a proliferation-required gene or its gene product lies
CC   or a gene on which the test compound that inhibits proliferation of an
CC   organism acts; (9) manufacturing an antibiotic; (10) profiling a
CC   compound's activity; (11) a culture comprising strains in which the gene
CC   product is overexpressed or underexpressed; (12) determining the extent
CC   to which each of the strains is present in a culture or collection of
CC   strains; or (13) identifying the target of a compound that inhibits the
CC   proliferation of an organism. The antisense nucleic acids are useful for
CC   identifying proteins or screening for homologous nucleic acids required
CC   for cellular proliferation to isolate candidate molecules for rational
CC   drug discovery programs, or for screening homologous nucleic acids
CC   required for proliferation in cells other than S. aureus, S. typhimurium,
CC   K. pneumoniae or P. aeruginosa. The present sequence is encoded by one of
CC   the target prokaryotic essential genes. Note: The sequence data for this
CC   patent did not form part of the printed specification, but was obtained
CC   in electronic format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 411 AA;

  Query Match             31.1%;  Score 57;  DB 6;  Length 411;
  Best Local Similarity   42.9%;  Pred. No. 31;
  Matches   12;  Conservative    6;  Mismatches   10;  Indels    0;  Gaps    0;

Qy          3 LSQEQLEHRERSLQTLRDIQRMLFPDEK 30
              |: |||:  | |||  | |  ::  |  |::
Db        207 LADEQLQQLEASLQALTDEEKRLLADQQ 234
```

```
RESULT 24
ABU47461
ID   ABU47461 standard; protein; 1034 AA.
XX
AC   ABU47461;
XX
DT   19-JUN-2003  (first entry)
XX
DE   Protein encoded by Prokaryotic essential gene #32988.
XX
KW   Antisense; prokaryotic essential gene; cell proliferation; drug design.
XX
OS   Salmomella typhi.
XX
PN   WO200277183-A2.
XX
PD   03-OCT-2002.
XX
PF   21-MAR-2002; 2002WO-US009107.
XX
PR   21-MAR-2001; 2001US-00815242.
PR   06-SEP-2001; 2001US-00948993.
PR   25-OCT-2001; 2001US-0342923P.
PR   08-FEB-2002; 2002US-00072851.
PR   06-MAR-2002; 2002US-0362699P.
XX
PA   (ELIT-) ELITRA PHARM INC.
XX
PI   Wang L,  Zamudio C,  Malone C,  Haselbeck R,  Ohlsen KL,  Zyskind JW;
PI   Wall D,  Trawick JD,  Carr GJ,  Yamamoto R,  Forsyth RA,  Xu HH;
XX
DR   WPI; 2003-029926/02.
DR   N-PSDB; ACA51331.
XX
PT   New antisense nucleic acids, useful for identifying proteins or screening
PT   for homologous nucleic acids required for cellular proliferation to
PT   isolate candidate molecules for rational drug discovery programs.
XX
PS   Claim 25; SEQ ID NO 75385; 1766pp; English.
XX
CC   The invention relates to an isolated nucleic acid comprising any one of
CC   the 6213 antisense sequences given in the specification where expression
CC   of the nucleic acid inhibits proliferation of a cell. Also included are:
CC   (1) a vector comprising a promoter operably linked to the nucleic acid
CC   encoding a polypeptide whose expression is inhibited by the antisense
CC   nucleic acid; (2) a host cell containing the vector; (3) an isolated
CC   polypeptide or its fragment whose expression is inhibited by the
CC   antisense nucleic acid; (4) an antibody capable of specifically binding
CC   the polypeptide;(5) producing the polypeptide; (6) inhibiting cellular
CC   proliferation or the activity of a gene in an operon required for
CC   proliferation; (7) identifying a compound that influences the activity of
CC   the gene product or that has an activity against a biological pathway
CC   required for proliferation, or that inhibits cellular proliferation; (8)
CC   identifying a gene required for cellular proliferation or the biological
CC   pathway in which a proliferation-required gene or its gene product lies
CC   or a gene on which the test compound that inhibits proliferation of an
CC   organism acts; (9) manufacturing an antibiotic; (10) profiling a
CC   compound's activity; (11) a culture comprising strains in which the gene
CC   product is overexpressed or underexpressed; (12) determining the extent
CC   to which each of the strains is present in a culture or collection of
CC   strains; or (13) identifying the target of a compound that inhibits the
CC   proliferation of an organism. The antisense nucleic acids are useful for
CC   identifying proteins or screening for homologous nucleic acids required
CC   for cellular proliferation to isolate candidate molecules for rational
CC   drug discovery programs, or for screening homologous nucleic acids
CC   required for proliferation in cells other than S. aureus, S. typhimurium,
CC   K. pneumoniae or P. aeruginosa. The present sequence is encoded by one of
CC   the target prokaryotic essential genes. Note: The sequence data for this
CC   patent did not form part of the printed specification, but was obtained
CC   in electronic format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 1034 AA;
```

```
  Query Match             31.1%;  Score 57;  DB 6;  Length 1034;
  Best Local Similarity   42.9%;  Pred. No. 86;
  Matches   12;  Conservative    6;  Mismatches   10;  Indels    0;  Gaps    0;

Qy          3 LSQEQLEHRERSLQTLRDIQRMLFPDEK 30
              |: |||:  | ||| | | :: |  |::
Db        213 LADEQLQQLEASLQALTDEEKRLLADQQ 240


RESULT 25
ADP04653
ID   ADP04653 standard; protein; 425 AA.
XX
AC   ADP04653;
XX
DT   29-JUL-2004  (first entry)
XX
DE   Sea squirt protein with tissue specific expression in development Seq248.
XX
KW   sea squirt; regeneration medicine; gene therapy; cell proliferation;
KW   differentiation; reproduction; environmental measurement; water survey.
XX
OS   Ciona intestinalis.
XX
PN   JP2004057129-A.
XX
PD   26-FEB-2004.
XX
PF   31-JUL-2002; 2002JP-00222593.
XX
PR   31-JUL-2002; 2002JP-00222593.
XX
PA   (KAGA-) KAGAKU GIJUTSU SHINKO JIGYODAN.
XX
DR   WPI; 2004-287079/27.
DR   N-PSDB; ADP04652.
XX
PT   Novel gene cluster which is specifically expressed in tissue or organ
PT   during developmental phase of sea squirt, useful for elucidation of
PT   mechanism of development of tissue or organ of sea squirt.
XX
PS   Claim 1; SEQ ID NO 248; 1846pp; Japanese.
XX
CC   This invention relates to novel genes and the encoded proteins thereof
CC   that are derived from the sea squirt Ciona intestinalis. Specifically, it
CC   refers to those genes that are expressed in the tissues or organs of the
CC   sea squirt during its developmental phase. The present invention
CC   describes the identification of these genes as useful for elucidation of
CC   the mechanism of development and hence for developing regeneration
CC   medicines and gene therapy techniques. Accordingly, they can be used in
CC   the research of various genetic diseases, as well as the analysis of cell
CC   proliferation, differentiation and reproduction. Furthermore, such
CC   compositions can be useful for environmental measurements and water
CC   surveys, particularly for sea water surveys, and also for the preparation
CC   of transformed sea squirt for improving edibility of sea squirt such as
CC   Halocynthia roretzi. This polypeptide sequence is a sea squirt protein
CC   sequence that has tissue specific expression during development, given in
CC   an exemplification of the invention.
XX
SQ   Sequence 425 AA;

  Query Match             30.6%;  Score 56;  DB 8;  Length 425;
  Best Local Similarity   36.4%;  Pred. No. 44;
  Matches   12;  Conservative    9;  Mismatches    8;  Indels    4;  Gaps    1;

Qy          3 LSQEQLEHRERSLQTLR----DIQRMLFPDEKE 31
              ::||::  :|| ||: |     |:|| : |: |
Db        391 MAQEEILRKERELQSARQKLAQIRRMRYKDDSE 423


RESULT 26
ABR53351
ID   ABR53351 standard; protein; 584 AA.
XX
AC   ABR53351;
XX
DT   20-JUN-2003  (first entry)
XX
DE   Protein sequence #SEQ ID 1567.
XX
KW   Multiprotein complex; eukaryote; drug target; diagnosis.
XX
OS   Saccharomyces cerevisiae.
XX
PN   EP1258494-A1.
XX
PD   20-NOV-2002.
XX
PF   20-DEC-2001; 2001EP-00130253.
XX
PR   15-MAY-2001; 2001EP-00111774.
XX
PA   (CELL-) CELLZOME AG.
XX
PI   Bauer A,  Gavin A,  Grandi P,  Krause R,  Kruse UD,  Kuester BD;
PI   Marzioch M,  Schultz JD,  Superti-Furga GD;
XX
DR   WPI; 2003-250078/25.
DR   N-PSDB; ACC61393.
XX
PT   New isolated protein complexes useful for diagnosing a disease or
PT   disorder, or as a target for an active agent of a pharmaceutical,
PT   preferably a drug target in the treatment or prevention of disease or
PT   disorder.
XX
PS   Disclosure; SEQ ID NO 1567; 17pp + Sequence Listing; English.
XX
CC   The invention relates to multiprotein complexes from eukaryotes. Proteins
CC   of the invention and DNA sequences encoding them are given in records
CC   ABR52568-ABR53903 and ACC60610-ACC61944 respectively. The complexes are
CC   obtainable by using a protein as a bait and isolating the set of proteins
CC   which is attached thereto from cells. Such protein complexes may comprise
CC   up to 30 distinct proteins. Protein complexes of the invention are useful
CC   for diagnosing a disease or disorder, or as a target for an active agent
CC   of a pharmaceutical, preferably a drug target in the treatment or
CC   prevention of a disease or disorder. Note: The sequence data for this
CC   patent is not represented in the printed specification, but is based on
CC   sequence information supplied by the European Patent Office. The complete
CC   document is available on CD-ROM
XX
SQ   Sequence 584 AA;

  Query Match             30.6%;  Score 56;  DB 6;  Length 584;
  Best Local Similarity   36.0%;  Pred. No. 63;
  Matches    9;  Conservative   10;  Mismatches    6;  Indels    0;  Gaps    0;

Qy          4 SQEQLEHRERSLQTLRDIQRMLFPD 28
              |:|::| | ::: |:|:  : |:||
Db        193 SREEIEIRNQTISTIREAVKQLWPD 217


RESULT 27
ADK63408
ID   ADK63408 standard; protein; 584 AA.
XX
AC   ADK63408;
XX
DT   06-MAY-2004  (first entry)
XX
DE   Disease treating protein complex-derived protein #949.
XX
KW   protein complex; drug target; diagnosis.
XX
OS   Unidentified.
XX
```

PN EP1338608-A2.
XX
PD 27-AUG-2003.
XX
PF 20-DEC-2002; 2002EP-00102902.
XX
PR 20-DEC-2001; 2001EP-00130253.
XX
PA (CELL-) CELLZOME AG.
XX
PI Bauer A, Gavin A, Superti-Furga G, Kuester B, Schultz J;
PI Marzioch M, Grandi P, Krause R, Kruse U, Merino A, Bauch A;
PI Michon A, Leutwein C, Rick J;
XX
DR WPI; 2003-638460/61.
DR N-PSDB; ADK63409.
XX
PT New proteins and protein complexes from eukaryotes, useful as targets in
PT drug screening, or in diagnosing or screening for the presence of a
PT disease or disorder, or a predisposition for developing a disease or
PT disorder in a subject.
XX
PS Disclosure; SEQ ID NO 1897; 13pp; English.
XX
CC The invention relates to novel protein complexes comprising a first and a
CC second protein, or its derivative, fragment, homologue or variant. The
CC proteins are selected from given protein complexes, which are not defined
CC in the specification. The variants are encoded by nucleic acids that
CC hybridize to the nucleic acids encoding the proteins under low stringency
CC conditions. The protein complexes are useful as targets for an active
CC agent of a pharmaceutical. These protein complexes are particularly
CC useful as drugs targets for the treatment or preventing of a disease or
CC disorder. The complexes and methods above are useful in diagnosing or
CC screening for the presence of a disease or disorder or a predisposition
CC for developing a disease or disorder in a subject. These are also useful
CC in screening for a drug for treatment or prevention of a disease or
CC disorder. The molecule that modulates the amount, activity or protein
CC components of the complex is useful for the manufacture of a medicament
CC for the treatment or prevention of a disease or disorder. This sequence
CC corresponds to a protein of the invention. (Note: the sequence data for
CC this patent did not form part of the printed specification but was
CC obtained from the EPO in electronic format).
XX
SQ Sequence 584 AA;

Query Match 30.6%; Score 56; DB 7; Length 584;
Best Local Similarity 36.0%; Pred. No. 63;
Matches 9; Conservative 10; Mismatches 6; Indels 0; Gaps 0;

Qy 4 SQEQLEHRERSLQTLRDIQRMLFPD 28
|:|::| | ::: |:|: : |:||
Db 193 SREEIEIRNQTISTIREAVKQLWPD 217

RESULT 28
ADN19362
ID ADN19362 standard; protein; 584 AA.
XX
AC ADN19362;
XX
DT 02-DEC-2004 (first entry)
XX
DE Bacterial polypeptide #2015.
XX
KW Recombinant DNA construct; transformed plant; improved plant property;
KW cold tolerance; heat tolerance; drought tolerance; herbicide; osmosis;
KW pathogen tolerance; pest tolerance; plant disease resistance;
KW cell cycle pathway modification; plant growth regulator;
KW homologous recombination; seed oil yield; protein yield; carbohydrate;
KW nitrogen; phosphorus; photosynthesis; lignin; galactomannan;
KW bacterial polypeptide.
XX
OS Bacteria.
XX
PN US2003233675-A1.
XX
PD 18-DEC-2003.
XX
PF 20-FEB-2003; 2003US-00369493.
XX
PR 21-FEB-2002; 2002US-0360039P.
XX
PA (CAOY/) CAO Y.
PA (HINK/) HINKLE G J.
PA (SLAT/) SLATER S C.
PA (CHEN/) CHEN X.
PA (GOLD/) GOLDMAN B S.
XX
PI Cao Y, Hinkle GJ, Slater SC, Chen X, Goldman BS;
XX
DR WPI; 2004-061375/06.
XX
PT New recombinant DNA construct comprising a promoter positioned to provide
PT for expression of a polynucleotide encoding a polypeptide from a
PT microbial source, useful for producing plants with improved properties.
XX
PS Claim 1; SEQ ID NO 2015; 122pp; English.
XX
CC The invention relates to a recombinant DNA construct comprising a
CC promoter functional in a plant cell, where the promoter is positioned to
CC provide for expression of a polynucleotide encoding a polypeptide from a
CC microbial source. The invention also relates to a transformed plant
CC comprising the recombinant DNA construct and a method of producing a
CC transformed plant having an improved property. The plant is a crop plant
CC such as maize or soybean. The method of producing a transformed plant
CC having an improved property comprises transforming a plant with the
CC recombinant DNA construct and growing the transformed plant, where the
CC polynucleotide or polypeptide is useful for improving plant properties.
CC The recombinant DNA construct is useful for producing plants with
CC improved plant properties, e.g. improved cold, heat or drought tolerance,
CC tolerance to herbicides, extreme osmotic conditions, pathogens or pests,
CC increased resistance to plant disease, better growth rate by modification
CC of the cell cycle pathway with plant growth regulators, increased rate of
CC homologous recombination, modified seed oil or protein yield and/or
CC content, improved yield by modification of carbohydrate, nitrogen or
CC phosphorus use and/or uptake, by modification of photosynthesis or by
CC providing improved plant growth and development under at least one stress
CC condition, improved lignin production or improved galactomannan
CC production. This sequence represents a bacterial polypeptide used in the
CC scope of the invention. Note: The sequence data for this patent did not
CC form part of the printed specification but was obtained in electronic
CC format from USPTO at seqdata.uspto.gov/sequence.html.
XX
SQ Sequence 584 AA;

Query Match 30.6%; Score 56; DB 8; Length 584;
Best Local Similarity 36.0%; Pred. No. 63;
Matches 9; Conservative 10; Mismatches 6; Indels 0; Gaps 0;

Qy 4 SQEQLEHRERSLQTLRDIQRMLFPD 28
|:|::| | ::: |:|: : |:||
Db 193 SREEIEIRNQTISTIREAVKQLWPD 217

RESULT 29
ADM26229
ID ADM26229 standard; protein; 818 AA.
XX
AC ADM26229;
XX
DT 20-MAY-2004 (first entry)
XX
DE Hyperthermophile Methanopyrus kandleri protein #835.
XX
KW hyperthermophile; protein stability enhancement;
KW protein activity enhancement.

XX
OS Methanopyrus kandleri.
XX
PN WO2003076575-A2.
XX
PD 18-SEP-2003.
XX
PF 04-MAR-2003; 2003WO-US006664.
XX
PR 04-MAR-2002; 2002US-0361742P.
PR 14-MAY-2002; 2002US-0380423P.
PR 16-SEP-2002; 2002US-0410974P.
XX
PA (FIDE-) FIDELITY SYSTEMS INC.
PA (MALY/) MALYKH A.
XX
PI Slesarev AI, Pavlov A, Pavlova N, Kozyavkin S;
XX
DR WPI; 2003-748383/70.
DR N-PSDB; ADM27081.
XX
PT New isolated nucleic acids encoding any of about 1700 Methanopyrus
PT kandleri proteins, and the encoded proteins, useful as a medicaments or
PT as diagnostic agents.
XX
PS Claim 31; SEQ ID NO 835; 1023pp; English.
XX
CC The invention comprises the amino acid sequence of proteins from the
CC hyperthermophile Methanopyrus kandleri, the invention also comprises the
CC complete genome from Methanopyrus kandleri. The Methanopyrus kandleri
CC proteins of the invention are useful for enhancing the stability and/or
CC activity of other proteins. The Methanopyrus kandleri genome is useful in
CC a variety of diagnostic and analytical methods. The present amino acid
CC sequence represents a Methanopyrus kandleri protein of the invention.
XX
SQ Sequence 818 AA;

Query Match 30.6%; Score 56; DB 7; Length 818;
Best Local Similarity 42.9%; Pred. No. 91;
Matches 15; Conservative 4; Mismatches 16; Indels 0; Gaps 0;

Qy 2 GLSQEQLEHRERSLQTLRDIQRMLFPDEKEFTGAQ 36
Db 134 GFSQTLLEKLERLLHELRDIDRVVEMVDPAFDPAE 168

RESULT 30
ADL83239
ID ADL83239 standard; protein; 1132 AA.
XX
AC ADL83239;
XX
DT 17-JUN-2004 (first entry)
XX
DE Human PRO84721, SEQ ID 441.
XX
KW Immunosuppressive; Cytostatic; Antiarthritic; Antirheumatic; Antianemic;
KW Antiallergic; Muscular; Neuroprotective; Nephrotropic; Antiinflammatory;
KW Gene Therapy; PRO; B cell related disorder; cancer;
KW immune-mediated inflammatory disease; human.
XX
OS Homo sapiens.
XX
PN WO2004024097-A2.
XX
PD 25-MAR-2004.
XX
PF 15-SEP-2003; 2003WO-US029097.
XX
PR 16-SEP-2002; 2002US-0411392P.
XX
PA (GETH) GENENTECH INC.
XX
PI Chiu H, Clark H, Dennis K, Fong S, Schoenfeld JR, Wood WI;
PI Wu TD;
XX
DR WPI; 2004-329389/30.
DR N-PSDB; ADL83238.
XX
PT New PRO polypeptide, useful for diagnosing and treating a B cell related
PT disorder, e.g. Burkitt's lymphoma, rheumatoid arthritis, autoimmune
PT mediated hemolytic anemia, myasthenia gravis or ankylosing spondylitis.
XX
PS Claim 10; Fig 441; 695pp; English.
XX
CC The present invention relates to PRO proteins and their coding sequences.
CC The PRO proteins are useful for diagnosing and treating a B cell related
CC disorder, e.g. X-linked infantile hypogammaglobulinemia, polysaccharide
CC antigen unresponsiveness, selective IgA deficiency, selective IgM
CC deficiency, selective deficiency of IgG subclasses, immunodeficiency with
CC hyper IgM, transient hypogammaglobulinemia of infancy, Burkitt's
CC lymphoma, intermediate lymphoma, follicular lymphoma, type II
CC hypersensitivity, rheumatoid arthritis, autoimmune mediated haemolytic
CC anaemia, myasthenia gravis, hypoadrenocorticism, glomerulonephritis, or
CC ankylosing spondylitis. The PRO proteins are also useful for preparing a
CC medicament for treating a condition that is responsive to the PRO
CC protein, e.g. cancer or immune-mediated inflammatory diseases. The PRO
CC coding sequences are useful as hybridization probes in chromosome and
CC gene mapping, in preparing PRO proteins, or in generating transgenic
CC animals or knockout animals, which in turn are useful in the development
CC and screening of therapeutically useful reagents.
XX
SQ Sequence 1132 AA;

Query Match 30.6%; Score 56; DB 8; Length 1132;
Best Local Similarity 33.3%; Pred. No. 1.3e+02;
Matches 10; Conservative 8; Mismatches 12; Indels 0; Gaps 0;

Qy 5 QEQLEHRERSLQTLRDIQRMLFPDEKEFTG 34
Db 590 ENQRSHQELISQLLQSYMKLLLPDDEKFHG 619

RESULT 31
ADQ17519
ID ADQ17519 standard; protein; 1132 AA.
XX
AC ADQ17519;
XX
DT 26-AUG-2004 (first entry)
XX
DE Human soft tissue sarcoma-upregulated protein - SEQ ID 336.
XX
KW soft tissue sarcoma; cytostatic; gene therapy; vaccine; screening; human.
XX
OS Homo sapiens.
XX
PN WO2004048938-A2.
XX
PD 10-JUN-2004.
XX
PF 26-NOV-2003; 2003WO-US038193.
XX
PR 26-NOV-2002; 2002US-0429739P.
XX
PA (PROT-) PROTEIN DESIGN LABS INC.
XX
PI Aziz N, Ginsburg WM, Zlotnik A;
XX
DR WPI; 2004-441208/41.
XX
PT Early detection of soft tissue sarcoma comprises determining expression
PT of a gene in a first soft tissue sample and a normal soft tissue sample
PT and comparing the gene expression, also useful in treating soft tissue
PT sarcoma.
XX

```
PS   Example 2; SEQ ID NO 336; 210pp; English.
XX
CC   The invention relates to a novel method for detecting soft tissue sarcoma
CC   which comprises obtaining a first soft tissue sample from an individual
CC   and a normal soft tissue sample from the same or different individual,
CC   determining the expression of a gene in both samples and comparing the
CC   expression of the gene in both soft tissue samples, where a higher level
CC   of protein expression in the first soft tissue sample indicates the
CC   presence of soft tissue sarcoma. The method of the invention has
CC   cytostatic applications and may be useful for detecting soft tissue
CC   sarcoma, possibly via gene therapy or vaccine production. The nucleic
CC   acid sequences may be useful in diagnostic and screening applications.
CC   The current sequence is that of a human soft tissue sarcoma-upregulated
CC   protein of the invention. The current sequence is not shown within the
CC   specification per se but was submitted in CD format by the inventor.
XX
SQ   Sequence 1132 AA;

  Query Match             30.6%;  Score 56;  DB 8;  Length 1132;
  Best Local Similarity   33.3%;  Pred. No. 1.3e+02;
  Matches   10;  Conservative    8;  Mismatches   12;  Indels    0;  Gaps    0;

Qy          5 QEQLEHRERSLQTLRDIQRMLFPDEKEFTG 34
              : |  |:|   | |:   ::| ||:::| |
Db        590 ENQRSHQELISQLLQSYMKLLLPDDEKFHG 619


RESULT 32
ABB63502
ID   ABB63502 standard; protein; 1294 AA.
XX
AC   ABB63502;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Drosophila melanogaster polypeptide SEQ ID NO 17298.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US009231.
XX
PR   23-MAR-2000; 2000US-0191637P.
PR   11-JUL-2000; 2000US-00614150.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Venter JC,  Adams M,  Li PWD,  Myers EW;
XX
DR   WPI; 2001-656860/75.
DR   N-PSDB; ABL07605.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signaling and cell-cell
PT   interactions.
XX
PS   Disclosure; SEQ ID NO 17298; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-
CC   ABB72072). The sequence data for this patent did not form part of the
CC   printed specification, but was obtained in electronic format directly
CC   from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 1294 AA;

  Query Match             30.6%;  Score 56;  DB 4;  Length 1294;
  Best Local Similarity   45.5%;  Pred. No. 1.5e+02;
  Matches   10;  Conservative    6;  Mismatches    6;  Indels    0;  Gaps    0;

Qy          3 LSQEQLEHRERSLQTLRDIQRM 24
              | |:| | :|:   ||||::|:
Db        897 LQQQQAESQEQQASTLRDLERL 918


RESULT 33
ABU17570
ID   ABU17570 standard; protein; 757 AA.
XX
AC   ABU17570;
XX
DT   19-JUN-2003  (first entry)
XX
DE   Protein encoded by Prokaryotic essential gene #3097.
XX
KW   Antisense; prokaryotic essential gene; cell proliferation; drug design.
XX
OS   Bacillus anthracis.
XX
PN   WO200277183-A2.
XX
PD   03-OCT-2002.
XX
PF   21-MAR-2002; 2002WO-US009107.
XX
PR   21-MAR-2001; 2001US-00815242.
PR   06-SEP-2001; 2001US-00948993.
PR   25-OCT-2001; 2001US-0342923P.
PR   08-FEB-2002; 2002US-00072851.
PR   06-MAR-2002; 2002US-0362699P.
XX
PA   (ELIT-) ELITRA PHARM INC.
XX
PI   Wang L,  Zamudio C,  Malone C,  Haselbeck R,  Ohlsen KL,  Zyskind JW;
PI   Wall D,  Trawick JD,  Carr GJ,  Yamamoto R,  Forsyth RA,  Xu HH;
XX
DR   WPI; 2003-029926/02.
DR   N-PSDB; ACA21440.
XX
PT   New antisense nucleic acids, useful for identifying proteins or screening
PT   for homologous nucleic acids required for cellular proliferation to
PT   isolate candidate molecules for rational drug discovery programs.
XX
PS   Claim 25; SEQ ID NO 45494; 1766pp; English.
XX
CC   The invention relates to an isolated nucleic acid comprising any one of
CC   the 6213 antisense sequences given in the specification where expression
CC   of the nucleic acid inhibits proliferation of a cell. Also included are:
CC   (1) a vector comprising a promoter operably linked to the nucleic acid
CC   encoding a polypeptide whose expression is inhibited by the antisense
CC   nucleic acid; (2) a host cell containing the vector; (3) an isolated
CC   polypeptide or its fragment whose expression is inhibited by the
CC   antisense nucleic acid; (4) an antibody capable of specifically binding
CC   the polypeptide;(5) producing the polypeptide; (6) inhibiting cellular
CC   proliferation or the activity of a gene in an operon required for
CC   proliferation; (7) identifying a compound that influences the activity of
CC   the gene product or that has an activity against a biological pathway
CC   required for proliferation, or that inhibits cellular proliferation; (8)
CC   identifying a gene required for cellular proliferation or the biological
CC   pathway in which a proliferation-required gene or its gene product lies
CC   or a gene on which the test compound that inhibits proliferation of an
CC   organism acts; (9) manufacturing an antibiotic; (10) profiling a
CC   compound's activity; (11) a culture comprising strains in which the gene
CC   product is overexpressed or underexpressed; (12) determining the extent
CC   to which each of the strains is present in a culture or collection of
```

```
CC   strains; or (13) identifying the target of a compound that inhibits the
CC   proliferation of an organism. The antisense nucleic acids are useful for
CC   identifying proteins or screening for homologous nucleic acids required
CC   for cellular proliferation to isolate candidate molecules for rational
CC   drug discovery programs, or for screening homologous nucleic acids
CC   required for proliferation in cells other than S. aureus, S. typhimurium,
CC   K. pneumoniae or P. aeruginosa. The present sequence is encoded by one of
CC   the target prokaryotic essential genes. Note: The sequence data for this
CC   patent did not form part of the printed specification, but was obtained
CC   in electronic format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 757 AA;

  Query Match             30.3%;  Score 55.5;  DB 6;  Length 757;
  Best Local Similarity   38.7%;  Pred. No. 98;
  Matches   12;  Conservative    8;  Mismatches    8;  Indels    3;  Gaps    1;

Qy          3 LSQEQLEHRERSLQTLRDIQRMLFPDEKEFT 33
              ::||||| :|   : :| ||: :  ||:  |
Db        180 VAQEQLEEQE---ENIRQIQKQMLADEERNT 207


RESULT 34
ADP56607
ID   ADP56607 standard; protein; 237 AA.
XX
AC   ADP56607;
XX
DT   09-SEP-2004  (first entry)
XX
DE   Human breast cancer-related protein - SEQ ID 141.
XX
KW   BSNA; BSP; cytostatic; breast cancer; vaccine; gene therapy; human;
KW   chromosome 22q13.31.
XX
OS   Homo sapiens.
XX
PN   WO2004053075-A2.
XX
PD   24-JUN-2004.
XX
PF   05-DEC-2003; 2003WO-US038739.
XX
PR   05-DEC-2002; 2002US-0431097P.
PR   05-DEC-2002; 2002US-0431122P.
XX
PA   (DIAD-) DIADEXUS INC.
XX
PI   Macina RA,  Turner LR,  Sun Y;
XX
DR   WPI; 2004-468847/44.
DR   N-PSDB; ADP56516.
XX
PT   New breast specific nucleic acid molecules and polypeptides useful for
PT   diagnosing, preventing or treating breast cancer, for producing
PT   transgenic animals or cells, or for research purposes.
XX
PS   Claim 12; SEQ ID NO 141; 387pp; English.
XX
CC   The invention relates to a novel isolated breast specific nucleic acid
CC   (BSNA) molecule which comprises a nucleic acid sequence encoding any of
CC   the 107 breast specific protein (BSP) amino acid sequences fully defined
CC   in the specification. The molecules of the invention demonstrate
CC   cytostatic activity and may be useful for diagnosing, preventing or
CC   treating breast cancer, possibly via vaccine production or gene therapy.
CC   The current sequence is that of a human breast cancer-related protein of
CC   the invention.
XX
SQ   Sequence 237 AA;

  Query Match             30.1%;  Score 55;  DB 8;  Length 237;
  Best Local Similarity   52.2%;  Pred. No. 32;
  Matches   12;  Conservative    4;  Mismatches    7;  Indels    0;  Gaps    0;

Qy          2 GLSQEQLEHRERSLQTLRDIQRM 24
              :|| | | |  |:||:|:|||:
Db         28 GLRTEGLFRRSASVQTVREIQRL 50


RESULT 35
ADP56608
ID   ADP56608 standard; protein; 248 AA.
XX
AC   ADP56608;
XX
DT   09-SEP-2004  (first entry)
XX
DE   Human breast cancer-related protein - SEQ ID 142.
XX
KW   BSNA; BSP; cytostatic; breast cancer; vaccine; gene therapy; human;
KW   chromosome 22q13.31.
XX
OS   Homo sapiens.
XX
PN   WO2004053075-A2.
XX
PD   24-JUN-2004.
XX
PF   05-DEC-2003; 2003WO-US038739.
XX
PR   05-DEC-2002; 2002US-0431097P.
PR   05-DEC-2002; 2002US-0431122P.
XX
PA   (DIAD-) DIADEXUS INC.
XX
PI   Macina RA,  Turner LR,  Sun Y;
XX
DR   WPI; 2004-468847/44.
DR   N-PSDB; ADP56516.
XX
PT   New breast specific nucleic acid molecules and polypeptides useful for
PT   diagnosing, preventing or treating breast cancer, for producing
PT   transgenic animals or cells, or for research purposes.
XX
PS   Claim 12; SEQ ID NO 142; 387pp; English.
XX
CC   The invention relates to a novel isolated breast specific nucleic acid
CC   (BSNA) molecule which comprises a nucleic acid sequence encoding any of
CC   the 107 breast specific protein (BSP) amino acid sequences fully defined
CC   in the specification. The molecules of the invention demonstrate
CC   cytostatic activity and may be useful for diagnosing, preventing or
CC   treating breast cancer, possibly via vaccine production or gene therapy.
CC   The current sequence is that of a human breast cancer-related protein of
CC   the invention.
XX
SQ   Sequence 248 AA;

  Query Match             30.1%;  Score 55;  DB 8;  Length 248;
  Best Local Similarity   52.2%;  Pred. No. 34;
  Matches   12;  Conservative    4;  Mismatches    7;  Indels    0;  Gaps    0;

Qy          2 GLSQEQLEHRERSLQTLRDIQRM 24
              || | |  |  |:||:|:|||:
Db         39 GLRTEGLFRRSASVQTVREIQRL 61


RESULT 36
AAB95073
ID   AAB95073 standard; protein; 294 AA.
XX
AC   AAB95073;
XX
DT   26-JUN-2001  (first entry)
XX
DE   Human protein sequence SEQ ID NO:16943.
```

```
XX
KW   Human; primer; detection; diagnosis; antisense therapy; gene therapy.
XX
OS   Homo sapiens.
XX
PN   EP1074617-A2.
XX
PD   07-FEB-2001.
XX
PF   28-JUL-2000; 2000EP-00116126.
XX
PR   29-JUL-1999;   99JP-00248036.
PR   27-AUG-1999;   99JP-00300253.
PR   11-JAN-2000; 2000JP-00118776.
PR   02-MAY-2000; 2000JP-00183767.
PR   09-JUN-2000; 2000JP-00241899.
XX
PA   (HELI-) HELIX RES INST.
XX
PI   Ota T,  Isogai T,  Nishikawa T,  Hayashi K,  Saito K,  Yamamoto J;
PI   Ishii S,  Sugiyama T,  Wakamatsu A,  Nagai K,  Otsuki T;
XX
DR   WPI; 2001-318749/34.
XX
PT   Primer sets for synthesizing polynucleotides, particularly the 5602 full-
PT   length cDNAs defined in the specification, and for the detection and/or
PT   diagnosis of the abnormality of the proteins encoded by the full-length
PT   cDNAs.
XX
PS   Claim 8; SEQ ID NO 16943; 2537pp + Sequence Listing; English.
XX
CC   The present invention describes primer sets for synthesising 5602 full-
CC   length cDNAs defined in the specification. Where a primer set comprises:
CC   (a) an oligo-dT primer and an oligonucleotide complementary to the
CC   complementary strand of a polynucleotide which comprises one of the 5602
CC   nucleotide sequences defined in the specification, where the
CC   oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC   of an oligonucleotide comprising a sequence complementary to the
CC   complementary strand of a polynucleotide which comprises a 5'-end
CC   sequence and an oligonucleotide comprising a sequence complementary to a
CC   polynucleotide which comprises a 3'-end sequence, where the
CC   oligonucleotide comprises at least 15 nucleotides and the combination of
CC   the 5'-end sequence/3'-end sequence is selected from those defined in the
CC   specification. The primer sets can be used in antisense therapy and in
CC   gene therapy. The primers are useful for synthesising polynucleotides,
CC   particularly full-length cDNAs. The primers are also useful for the
CC   detection and/or diagnosis of the abnormality of the proteins encoded by
CC   the full-length cDNAs. The primers allow obtaining of the full-length
CC   cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC   AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to AAB95893
CC   represent human amino acid sequences; and AAH13629 to AAH13632 represent
CC   oligonucleotides, all of which are used in the exemplification of the
CC   present invention
XX
SQ   Sequence 294 AA;

  Query Match             30.1%;  Score 55;  DB 4;  Length 294;
  Best Local Similarity   52.2%;  Pred. No. 41;
  Matches   12;  Conservative    4;  Mismatches    7;  Indels    0;  Gaps    0;

Qy          2 GLSQEQLEHRERSLQTLRDIQRM 24
              ||  | |  |  |:||:|:|||:
Db        224 GLRTEGLFRRSASVQTVREIQRL 246


RESULT 37
ABR82444
ID   ABR82444 standard; protein; 294 AA.
XX
AC   ABR82444;
XX
DT   06-NOV-2003  (first entry)
XX
DE   Human ARP13 polypeptide.
XX
KW   ARP; prostate; neoplastic; androgen responsive prostate; ARP15;
KW   cytostatic; gene therapy; human; ARP13.
XX
OS   Homo sapiens.
XX
PN   WO2003060148-A2.
XX
PD   24-JUL-2003.
XX
PF   15-JAN-2003; 2003WO-US001457.
XX
PR   15-JAN-2002; 2002US-00053248.
XX
PA   (SYST-) INST SYSTEMS BIOLOGY.
XX
PI   Lin B;
XX
DR   WPI; 2003-587287/55.
DR   N-PSDB; ACF35892.
XX
PT   Diagnosing or predicting susceptibility to a prostate neoplastic
PT   condition by contacting a specimen from the individual with an ARP15
PT   binding agent that selectively binds an ARP15 polypeptide.
XX
PS   Claim 78; Page 195-197; 227pp; English.
XX
CC   The invention relates to diagnosing or predicting susceptibility to a
CC   prostate neoplastic condition. The method involves (a) contacting a
CC   specimen from the individual with an androgen responsive prostate
CC   specific (ARP)15 binding agent that selectively binds an ARP15
CC   polypeptide; (b) determining a test expression level of ARP15 polypeptide
CC   in the specimen; and (c) comparing the test expression level to a non-
CC   neoplastic control expression level of ARP15 polypeptide, where an
CC   altered test expression level as compared to the control expression level
CC   indicates the presence of a prostate neoplastic condition in the
CC   individual. The method is useful for diagnosing or predicting
CC   susceptibility to a prostate neoplastic condition or for treating or
CC   reducing severity of a prostate neoplastic condition. The present
CC   sequence represents a human ARP13 polypeptide
XX
SQ   Sequence 294 AA;

  Query Match             30.1%;  Score 55;  DB 6;  Length 294;
  Best Local Similarity   52.2%;  Pred. No. 41;
  Matches   12;  Conservative    4;  Mismatches    7;  Indels    0;  Gaps    0;

Qy          2 GLSQEQLEHRERSLQTLRDIQRM 24
              ||  | |  |  |:||:|:|||:
Db        224 GLRTEGLFRRSASVQTVREIQRL 246


RESULT 38
ADQ74859
ID   ADQ74859 standard; protein; 294 AA.
XX
AC   ADQ74859;
XX
DT   07-OCT-2004  (first entry)
XX
DE   Human androgen responsive prostate specific (ARP) polypeptide #6.
XX
KW   Human; androgen responsive prostate specific polypeptide; ARP;
KW   prostate neoplastic condition; prostate cancer; cytostatic.
XX
OS   Homo sapiens.
XX
PN   US2004137440-A1.
XX
PD   15-JUL-2004.
XX
PF   15-JAN-2003; 2003US-00345837.
```

```
XX
PR   15-JAN-2003; 2003US-00345837.
XX
PA   (LINB/) LIN B.
XX
PI   Lin B;
XX
DR   WPI; 2004-517182/49.
DR   N-PSDB; ADQ74858.
XX
PT   New substantially pure androgen responsive specific nucleic acid, useful
PT   for diagnosing and treating prostate cancer.
XX
PS   Claim 89; SEQ ID NO 12; 102pp; English.
XX
CC   The invention relates to human androgen responsive prostate specific
CC   (ARP) polynucleotides and the polypeptides they encode. The invention
CC   also relates to a method of diagnosing or predicting susceptibility to a
CC   prostate neoplastic condition in an individual and a method for treating
CC   or reducing the severity of a prostate neoplastic condition in an
CC   individual. The polynucleotides, polypeptides and methods of the
CC   invention are useful for diagnosing and treating prostate cancer. This
CC   sequence represents a human ARP polypeptide of the invention.
XX
SQ   Sequence 294 AA;

  Query Match             30.1%;  Score 55;  DB 8;  Length 294;
  Best Local Similarity   52.2%;  Pred. No. 41;
  Matches   12;  Conservative    4;  Mismatches    7;  Indels    0;  Gaps    0;

Qy          2 GLSQEQLEHRERSLQTLRDIQRM 24
              ||  | |  |  |:||:|:|||:
Db        224 GLRTEGLFRRSASVQTVREIQRL 246


RESULT 39
ADM83551
ID   ADM83551 standard; protein; 333 AA.
XX
AC   ADM83551;
XX
DT   03-JUN-2004  (first entry)
XX
DE   Human Rho GTPase activating protein 8.
XX
KW   GTPase-activating protein; GTPAP; cell signalling;
KW   cell proliferative disorder; colon cancer; immune disorder; cytostatic;
KW   human; Rho GTPase activating protein 8.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Region          176. .189
FT                   /note= "Proline-rich region"
XX
PN   US2003129655-A1.
XX
PD   10-JUL-2003.
XX
PF   29-OCT-2002; 2002US-00284753.
XX
PR   18-FEB-2000; 2000US-00507765.
XX
PA   (INCY-) INCYTE GENOMICS INC.
XX
PI   Klinger TM,  Stewart EA,  Yue H,  Baughn MR;
XX
DR   WPI; 2003-829559/77.
XX
PT   New GTPase-activating proteins designated GTPAP-1 and its variant GTPAP-2
PT   are useful to diagnose, stage, treat and monitor cell signaling and cell
PT   proliferative disorders, particularly colon cancer.
XX
PS   Disclosure; SEQ ID NO 32; 64pp; English.
XX
CC   The present invention relates to novel GTPase-activating proteins
CC   (collectively designated as GTPAP), GTPAP-1 or its variant GTPAP-2 and
CC   their encoding cDNAs. The protein is used to diagnose, stage, treat and
CC   monitor cell signalling, immune and cell proliferative disorders,
CC   particularly colon cancer. The present sequence is human Rho GTPase
CC   activating protein 8 used in the invention.
XX
SQ   Sequence 333 AA;

  Query Match             30.1%;  Score 55;  DB 7;  Length 333;
  Best Local Similarity   52.2%;  Pred. No. 47;
  Matches   12;  Conservative    4;  Mismatches    7;  Indels    0;  Gaps    0;

Qy          2 GLSQEQLEHRERSLQTLRDIQRM 24
              ||  | |  |  |:||:|:|||:
Db        124 GLRTEGLFRRSASVQTVREIQRL 146


RESULT 40
ADP56606
ID   ADP56606 standard; protein; 337 AA.
XX
AC   ADP56606;
XX
DT   09-SEP-2004  (first entry)
XX
DE   Human breast cancer-related protein - SEQ ID 140.
XX
KW   BSNA; BSP; cytostatic; breast cancer; vaccine; gene therapy; human;
KW   chromosome 22q13.31.
XX
OS   Homo sapiens.
XX
PN   WO2004053075-A2.
XX
PD   24-JUN-2004.
XX
PF   05-DEC-2003; 2003WO-US038739.
XX
PR   05-DEC-2002; 2002US-0431097P.
PR   05-DEC-2002; 2002US-0431122P.
XX
PA   (DIAD-) DIADEXUS INC.
XX
PI   Macina RA,  Turner LR,  Sun Y;
XX
DR   WPI; 2004-468847/44.
DR   N-PSDB; ADP56515.
XX
PT   New breast specific nucleic acid molecules and polypeptides useful for
PT   diagnosing, preventing or treating breast cancer, for producing
PT   transgenic animals or cells, or for research purposes.
XX
PS   Claim 12; SEQ ID NO 140; 387pp; English.
XX
CC   The invention relates to a novel isolated breast specific nucleic acid
CC   (BSNA) molecule which comprises a nucleic acid sequence encoding any of
CC   the 107 breast specific protein (BSP) amino acid sequences fully defined
CC   in the specification. The molecules of the invention demonstrate
CC   cytostatic activity and may be useful for diagnosing, preventing or
CC   treating breast cancer, possibly via vaccine production or gene therapy.
CC   The current sequence is that of a human breast cancer-related protein of
CC   the invention.
XX
SQ   Sequence 337 AA;

  Query Match             30.1%;  Score 55;  DB 8;  Length 337;
  Best Local Similarity   52.2%;  Pred. No. 47;
  Matches   12;  Conservative    4;  Mismatches    7;  Indels    0;  Gaps    0;

Qy          2 GLSQEQLEHRERSLQTLRDIQRM 24
```

```
                ||  | |  |  |:||:|:|||:
Db        128 GLRTEGLFRRSASVQTVREIQRL 150

Search completed: June  8, 2005, 03:17:49
Job time : 135.875 secs
```

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: June 8, 2005, 02:59:37 ; Search time 104.125 Seconds
(without alignments)
104.003 Million cell updates/sec

Title: US-09-915-543-15_COPY_177_204
Perfect score: 136
Sequence: 1 VYVFSTEMANKAAEAVLKGQVETIVSFH 28

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 2105692 seqs, 386760381 residues

Total number of hits satisfying chosen parameters: 2105692

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_16Dec04:*
1: geneseqp1980s:*
2: geneseqp1990s:*
3: geneseqp2000s:*
4: geneseqp2001s:*
5: geneseqp2002s:*
6: geneseqp2003as:*
7: geneseqp2003bs:*
8: geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	136	100.0	28	8	ADJ71891	Adj71891 Human Lgs
2	136	100.0	1394	8	ADQ18945	Adq18945 Human sof
3	136	100.0	1426	5	AAB71229	Aab71229 Human leg
4	136	100.0	1426	7	ABW01534	Abw01534 Human lgs
5	136	100.0	1426	7	ADJ70152	Adj70152 Human hea
6	136	100.0	1426	8	ADJ71903	Adj71903 Human Lgs
7	136	100.0	1435	4	ABB11808	Abb11808 Human BCL
8	129	94.9	140	4	AAO05855	Aao05855 Human pol
9	98	72.1	28	7	ABW01528	Abw01528 Drosophil
10	98	72.1	28	8	ADJ71890	Adj71890 Fruit fly
11	98	72.1	1429	4	ABB58779	Abb58779 Drosophil
12	98	72.1	1464	5	AAB71228	Aab71228 D. melano
13	98	72.1	1464	7	ABW01527	Abw01527 Drosophil
14	98	72.1	1464	8	ADJ71911	Adj71911 Fruit fly
15	94	69.1	1494	5	AAU78460	Aau78460 Mouse bet
16	91	66.9	114	5	ABP06595	Abp06595 Human ORF
17	67	49.3	320	5	AAU78461	Aau78461 Mouse bet
18	65	47.8	113	4	AAO07544	Aao07544 Human pol
19	53	39.0	133	8	ADN48110	Adn48110 Thermococ
20	53	39.0	322	7	ABO66692	Abo66692 Klebsiell
21	52	38.2	360	3	AAG33446	Aag33446 Zea mays
22	52	38.2	448	3	AAG33445	Aag33445 Zea mays
23	52	38.2	509	3	AAG33444	Aag33444 Zea mays
24	51	37.5	1049	4	ABB60387	Abb60387 Drosophil
25	50	36.8	130	5	ABB89793	Abb89793 Human pol
26	50	36.8	365	4	ABG15088	Abg15088 Novel hum
27	50	36.8	621	5	AAE15740	Aae15740 Human ami
28	50	36.8	631	3	AAB43285	Aab43285 Human ORF
29	50	36.8	694	8	ABM80775	Abm80775 Tumour-as
30	50	36.8	961	8	ADI16244	Adi16244 Human nuc
31	50	36.8	1063	5	ABB08919	Abb08919 Human ami
32	50	36.8	1063	7	ADJ70652	Adj70652 Human hea
33	50	36.8	1078	7	ADK40961	Adk40961 Novel hum
34	50	36.8	1078	8	ADR15680	Adr15680 Kinase 69
35	49	36.0	187	5	ABP66271	Abp66271 Bifidobac
36	49	36.0	265	7	ABO61084	Abo61084 Klebsiell
37	49	36.0	586	6	ABU49676	Abu49676 Protein e
38	49	36.0	984	6	ABJ25889	Abj25889 Aspergill
39	49	36.0	1058	6	ABJ26489	Abj26489 Aspergill
40	48	35.3	310	7	ADC50023	Adc50023 Gene repa
41	48	35.3	330	8	ADS28390	Ads28390 Bacterial
42	48	35.3	498	3	AAG49364	Aag49364 Arabidops
43	48	35.3	498	3	AAG17973	Aag17973 Arabidops
44	48	35.3	609	3	AAG49363	Aag49363 Arabidops
45	48	35.3	609	3	AAG17972	Aag17972 Arabidops

ALIGNMENTS

RESULT 1
ADJ71891
ID ADJ71891 standard; peptide; 28 AA.
XX
AC ADJ71891;
XX
DT 20-MAY-2004 (first entry)
XX
DE Human Lgs/Bc19 peptide fragment #1.
XX
KW Human; legless; lgs; cell differentiation disorder;
KW cell proliferation disorder; cancer; Wnt pathway; medulloblastoma; colon;
KW breast; head; neck; brain; thyroid; skin; blood disease;
KW tissue regeneration; tissue repair; cytostatic; Lgs/Bc19.
XX
OS Homo sapiens.
XX
PN US2004038901-A1.
XX
PD 26-FEB-2004.
XX
PF 22-SEP-2003; 2003US-00664859.
XX
PR 28-JUL-2000; 2000US-0221502P.
PR 27-JUL-2001; 2001US-00915543.
XX
PA (UYZU-) UNIV ZURICH.
XX
PI Basler K, Brunner E, Froesch B, Kramps T, Peter O;
XX
DR WPI; 2004-203288/19.
XX
PT Novel polypeptide sharing one or more homologue amino acid domains with
PT Legless protein being functional homologue of Legless, useful for
PT diagnosing disorders of cell fate.
XX
PS Disclosure; SEQ ID NO 3; 62pp; English.
XX
CC The invention relates to a polypeptide sharing one or more homologous
CC amino acid domains with a Legless (Lgs) protein and is therefore a
CC functional homologue of Lgs. The invention also relates to a nucleotide
CC sequence encoding a protein present in invertebrate and/or vertebrate
CC organisms, the nucleotide sequence encoding a protein comprising a
CC positive function in a regulatory pathway and the use of the polypeptide
CC for the isolation of Lgs-binding proteins by carrying out an assay chosen
CC from an in vitro binding assay with such a peptide or a co-
CC immunoprecipitation from vertebrate or invertebrate cell lysates or a
CC mammalian or yeast two hybrid assay. The polypeptide and polynucleotide

CC are useful for treating disorders of cell fate, which involves
CC administering therapeutic compounds chosen from invertebrate and
CC vertebrate Lgs protein homologues or fragments, antibodies, antibody
CC fragments, lgs antisense DNA, lgs antisense RNA, lgs double-stranded RNA,
CC small peptides or chemical and natural compounds being capable of
CC interfering with Lgs function, synthesis and degradation. The disorders
CC are related to cell differentiation or cell proliferation. The compound
CC is administered to treat a cancerous condition by preventing progression
CC from a pre-neoplastic or non-malignant condition to a neoplastic or
CC malignant state. The cancerous condition is characterised by over-
CC stimulation of the Wnt pathway and is medulloblastoma or cancer of the
CC colon, breast, head and neck, brain, thyroid or skin. The therapeutic
CC compound may also be administered to a blood disease to promote tissue
CC regeneration and repair. This sequence represents a human Lgs/Bc19
CC peptide fragment of the invention.
XX
SQ Sequence 28 AA;

Query Match 100.0%; Score 136; DB 8; Length 28;
Best Local Similarity 100.0%; Pred. No. 5.2e-14;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 VYVFSTEMANKAAEAVLKGQVETIVSFH 28
||||||||||||||||||||||||||||
Db 1 VYVFSTEMANKAAEAVLKGQVETIVSFH 28

RESULT 2
ADQ18945
ID ADQ18945 standard; protein; 1394 AA.
XX
AC ADQ18945;
XX
DT 26-AUG-2004 (first entry)
XX
DE Human soft tissue sarcoma-upregulated protein - SEQ ID 1764.
XX
KW soft tissue sarcoma; cytostatic; gene therapy; vaccine; screening; human.
XX
OS Homo sapiens.
XX
PN WO2004048938-A2.
XX
PD 10-JUN-2004.
XX
PF 26-NOV-2003; 2003WO-US038193.
XX
PR 26-NOV-2002; 2002US-0429739P.
XX
PA (PROT-) PROTEIN DESIGN LABS INC.
XX
PI Aziz N, Ginsburg WM, Zlotnik A;
XX
DR WPI; 2004-441208/41.
XX
PT Early detection of soft tissue sarcoma comprises determining expression
PT of a gene in a first soft tissue sample and a normal soft tissue sample
PT and comparing the gene expression, also useful in treating soft tissue
PT sarcoma.
XX
PS Example 2; SEQ ID NO 1764; 210pp; English.
XX
CC The invention relates to a novel method for detecting soft tissue sarcoma
CC which comprises obtaining a first soft tissue sample from an individual
CC and a normal soft tissue sample from the same or different individual,
CC determining the expression of a gene in both samples and comparing the
CC expression of the gene in both soft tissue samples, where a higher level
CC of protein expression in the first soft tissue sample indicates the
CC presence of soft tissue sarcoma. The method of the invention has
CC cytostatic applications and may be useful for detecting soft tissue
CC sarcoma, possibly via gene therapy or vaccine production. The nucleic
CC acid sequences may be useful in diagnostic and screening applications.
CC The current sequence is that of a human soft tissue sarcoma-upregulated
CC protein of the invention. The current sequence is not shown within the
CC specification per se but was submitted in CD format by the inventor.
XX
SQ Sequence 1394 AA;

Query Match 100.0%; Score 136; DB 8; Length 1394;
Best Local Similarity 100.0%; Pred. No. 5.5e-12;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 VYVFSTEMANKAAEAVLKGQVETIVSFH 28
||||||||||||||||||||||||||||
Db 177 VYVFSTEMANKAAEAVLKGQVETIVSFH 204

RESULT 3
AAB71229
ID AAB71229 standard; protein; 1426 AA.
XX
AC AAB71229;
XX
DT 18-NOV-2002 (first entry)
XX
DE Human legless homologue lgs/bc19 protein.
XX
KW Legless; human; lgs; Wnt/Wingless signaling pathway; Wnt; Wg;
KW tissue proliferation; tumour; cytostatic; cellular disorder; colon;
KW blood disorder; cancer; breast; head and neck cancer; brain; thyroid;
KW medulloblastoma; skin cancer; tissue regeneration; tissue repair.
XX
OS Homo sapiens.
XX
PN US2002086986-A1.
XX
PD 04-JUL-2002.
XX
PF 27-JUL-2001; 2001US-00915543.
XX
PR 28-JUL-2000; 2000US-0221502P.
XX
PA (BASL/) BASLER K.
PA (BRUN/) BRUNNER E.
PA (FROE/) FROESCH B.
PA (KRAM/) KRAMPS T.
PA (PETE/) PETER O.
XX
PI Basler K, Brunner E, Froesch B, Kramps T, Peter O;
XX
DR WPI; 2002-635689/68.
DR N-PSDB; AAF88467.
XX
PT Novel polypeptide useful in therapeutic method for treating disorders of
PT cell fate such as cell differentiation or cell proliferation.
XX
PS Example II; Fig 8B; 41pp; English.
XX
CC This invention describes a novel polypeptide sharing one or more
CC homologous amino acid domains with the legless (lgs) protein, a
CC downstream component of the Wnt/Wingless (Wnt/Wg) signaling pathway
CC involved in the formation and maintenance of spatial arrangements and
CC proliferation of tissues during development, and in the formation and
CC growth of many human tumours. The products of the invention have
CC cytostatic activity and can be used to treat cellular disorders, blood
CC disorders and cancers caused by over-stimulation of the Wnt pathway,
CC where the cancerous condition is colon, breast, head and neck, brain,
CC thyroid, medulloblastoma or skin cancer. The product could also be used
CC to promote tissue regeneration and repair. This sequence represents the
CC human legless (lgs) protein homologue lgs/bc19 described in the
CC disclosure of the invention
XX
SQ Sequence 1426 AA;

Query Match 100.0%; Score 136; DB 5; Length 1426;
Best Local Similarity 100.0%; Pred. No. 5.7e-12;

```
  Matches   28;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 VYVFSTEMANKAAEAVLKGQVETIVSFH 28
              ||||||||||||||||||||||||||||
Db        177 VYVFSTEMANKAAEAVLKGQVETIVSFH 204


RESULT 4
ABW01534
ID   ABW01534 standard; protein; 1426 AA.
XX
AC   ABW01534;
XX
DT   15-JAN-2004  (first entry)
XX
DE   Human lgs/bcl9 protein.
XX
KW   Legless protein; lgs; cell fate disorder; blood disease; gene therapy;
KW   cancer; tissue regeneration; tissue repair; cytostatic.
XX
OS   Homo sapiens.
XX
PN   US2003114413-A1.
XX
PD   19-JUN-2003.
XX
PF   19-DEC-2002; 2002US-00322579.
XX
PR   28-JUL-2000; 2000US-0221502P.
PR   27-JUL-2001; 2001US-00915543.
XX
PA   (UYZU-) UNIV ZURICH.
XX
PI   Basler K,  Brunner E,  Froesch B,  Kramps T,  Peter O;
XX
DR   WPI; 2003-829432/77.
DR   N-PSDB; AAD62642.
XX
PT   Novel Lgs polypeptide useful for isolation of Lgs-binding proteins,
PT   diagnosing disorders of cell fate, treating diseases such as cancer.
XX
PS   Example 2; Fig 8B; 0pp; English.
XX
CC   The invention relates to novel legless (lgs) proteins and polynucleotides
CC   encoding such proteins. Lgs sequences are useful for the treatment of
CC   disorders of cell fate such as differentiation or proliferation. The
CC   invention is used to treat blood disease or a cancerous condition
CC   characterised by over-stimulation of the Wnt pathway such as colon,
CC   breast, head and neck, brain, thyroid, medulloblastoma or skin cancer and
CC   is administered to prevent progression from a pre-neoplastic or non-
CC   malignant condition to a neoplastic or malignant state. It is
CC   administered to promote tissue regeneration and repair. The invention is
CC   also useful in the therapy of diseases cost by an over-activation of Wg
CC   pathway. It is useful for reducing lgs gene expression in an invertebrate
CC   or vertebrate organism or an invertebrate or vertebrate cell line. The
CC   invention is also useful in gene therapy. The present sequence is human
CC   lgs/bcl9 protein used in the invention
XX
SQ   Sequence 1426 AA;

  Query Match             100.0%;  Score 136;  DB 7;  Length 1426;
  Best Local Similarity   100.0%;  Pred. No. 5.7e-12;
  Matches   28;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 VYVFSTEMANKAAEAVLKGQVETIVSFH 28
              ||||||||||||||||||||||||||||
Db        177 VYVFSTEMANKAAEAVLKGQVETIVSFH 204


RESULT 5
ADJ70152
ID   ADJ70152 standard; protein; 1426 AA.
XX
AC   ADJ70152;
XX
DT   06-MAY-2004  (first entry)
XX
DE   Human heat mitochondrial protein as a therapeutic target SeqID1958.
XX
KW   mitochondrial; human; screening assay; diabetes mellitus;
KW   Huntington's disease; osteoarthritis;
KW   Leber's hereditary optic neuropathy; LHON;
KW   mitochondrial encephalopathy lactic acidosis and stroke; MELAS;
KW   myoclonic epilepsy ragged red fibre syndrome; MERRF; cancer;
KW   neuroprotective; nootropic; antidiabetic; anticonvulsant; antiarthritic;
KW   osteopathic; ophthalmological; cytostatic.
XX
OS   Homo sapiens.
XX
PN   WO2003087768-A2.
XX
PD   23-OCT-2003.
XX
PF   04-APR-2003; 2003WO-US010870.
XX
PR   12-APR-2002; 2002US-0372843P.
PR   17-JUN-2002; 2002US-0389987P.
PR   20-SEP-2002; 2002US-0412418P.
XX
PA   (MITO-) MITOKOR.
PA   (BUCK-) BUCK INST AGE RES.
XX
PI   Ghosh SS,  Fahy ED,  Zhang B,  Gibson BW,  Taylor SW,  Glenn GM;
PI   Warnock DE;
XX
DR   WPI; 2003-845369/78.
XX
PT   Identifying a mitochondrial target for drug screening assays and for
PT   treating diseases associated with altered mitochondrial function,
PT   comprises detecting a modified polypeptide in a sample and correlating
PT   with the disease.
XX
PS   Claim 1; SEQ ID NO 1958; 180pp; English.
XX
CC   This invention relates to novel mitochondrial targets that can be used
CC   for therapeutic intervention in treating a disease associated with
CC   altered mitochondrial function. Specifically, it refers to a method for
CC   identifying proteins of the human heart mitochondrial proteome that are
CC   useful for drug screening assays, as well as therapeutic targets. The
CC   present invention describes a method for identifying such proteins that
CC   can be used in the treatment of various diseases associated with altered
CC   mitochondrial function including diabetes mellitus, Huntington's disease,
CC   osteoarthritis, Leber's hereditary optic neuropathy (LHON), mitochondrial
CC   encephalopathy lactic acidosis and stroke (MELAS), myoclonic epilepsy
CC   ragged red fibre syndrome (MERRF) or cancer. Accordingly, these
CC   compositions have neuroprotective, nootropic, antidiabetic,
CC   anticonvulsant, antiarthritic, osteopathic, ophthalmological and
CC   cytostatic activities. This polypeptide sequence is a human heart
CC   mitochondrial protein of the invention.
XX
SQ   Sequence 1426 AA;

  Query Match             100.0%;  Score 136;  DB 7;  Length 1426;
  Best Local Similarity   100.0%;  Pred. No. 5.7e-12;
  Matches   28;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 VYVFSTEMANKAAEAVLKGQVETIVSFH 28
              ||||||||||||||||||||||||||||
Db        177 VYVFSTEMANKAAEAVLKGQVETIVSFH 204


RESULT 6
ADJ71903
ID   ADJ71903 standard; protein; 1426 AA.
XX
AC   ADJ71903;
```

XX
DT 20-MAY-2004 (first entry)
XX
DE Human Lgs/Bc19 polypeptide.
XX
KW Human; legless; lgs; cell differentiation disorder;
KW cell proliferation disorder; cancer; Wnt pathway; medulloblastoma; colon;
KW breast; head; neck; brain; thyroid; skin; blood disease;
KW tissue regeneration; tissue repair; cytostatic; Lgs/Bc19.
XX
OS Homo sapiens.
XX
PN US2004038901-A1.
XX
PD 26-FEB-2004.
XX
PF 22-SEP-2003; 2003US-00664859.
XX
PR 28-JUL-2000; 2000US-0221502P.
PR 27-JUL-2001; 2001US-00915543.
XX
PA (UYZU-) UNIV ZURICH.
XX
PI Basler K, Brunner E, Froesch B, Kramps T, Peter O;
XX
DR WPI; 2004-203288/19.
DR N-PSDB; ADJ71902.
XX
PT Novel polypeptide sharing one or more homologue amino acid domains with
PT Legless protein being functional homologue of Legless, useful for
PT diagnosing disorders of cell fate.
XX
PS Example 2; SEQ ID NO 15; 62pp; English.
XX
CC The invention relates to a polypeptide sharing one or more homologous
CC amino acid domains with a Legless (Lgs) protein and is therefore a
CC functional homologue of Lgs. The invention also relates to a nucleotide
CC sequence encoding a protein present in invertebrate and/or vertebrate
CC organisms, the nucleotide sequence encoding a protein comprising a
CC positive function in a regulatory pathway and the use of the polypeptide
CC for the isolation of Lgs-binding proteins by carrying out an assay chosen
CC from an in vitro binding assay with such a peptide or a co-
CC immunoprecipitation from vertebrate or invertebrate cell lysates or a
CC mammalian or yeast two hybrid assay. The polypeptide and polynucleotide
CC are useful for treating disorders of cell fate, which involves
CC administering therapeutic compounds chosen from invertebrate and
CC vertebrate Lgs protein homologues or fragments, antibodies, antibody
CC fragments, lgs antisense DNA, lgs antisense RNA, lgs double-stranded RNA,
CC small peptides or chemical and natural compounds being capable of
CC interfering with Lgs function, synthesis and degradation. The disorders
CC are related to cell differentiation or cell proliferation. The compound
CC is administered to treat a cancerous condition by preventing progression
CC from a pre-neoplastic or non-malignant condition to a neoplastic or
CC malignant state. The cancerous condition is characterised by over-
CC stimulation of the Wnt pathway and is medulloblastoma or cancer of the
CC colon, breast, head and neck, brain, thyroid or skin. The therapeutic
CC compound may also be administered to a blood disease to promote tissue
CC regeneration and repair. This sequence represents the human Lgs/Bc19
CC polypeptide of the invention.
XX
SQ Sequence 1426 AA;

Query Match 100.0%; Score 136; DB 8; Length 1426;
Best Local Similarity 100.0%; Pred. No. 5.7e-12;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 VYVFSTEMANKAAEAVLKGQVETIVSFH 28
||||||||||||||||||||||||||||
Db 177 VYVFSTEMANKAAEAVLKGQVETIVSFH 204

RESULT 7
ABB11808
ID ABB11808 standard; peptide; 1435 AA.
XX
AC ABB11808;
XX
DT 11-JAN-2002 (first entry)
XX
DE Human BCL9 homologue, SEQ ID NO:2178.
XX
KW Human; cytokine; cell proliferation; cell differentiation; growth factor;
KW haematopoiesis regulation; tissue growth; immunomodulator; activin;
KW inhibin; chemotaxis; chemokinesis; thrombolysis; oncogenesis;
KW proliferation; metastasis; cancer; tumour; haematopoietic disorder;
KW myeloid cell disorder; lymphoid cell disorder; asthma; arthritis;
KW chronic inflammatory condition; proliferative retinopathy;
KW atherosclerosis; coronary heart disease; arterial ischaemia;
KW bone disorder; osteoporosis; vascular growth disorder;
KW tissue regeneration; wound healing; infection; immune disorder;
KW cell culture; drug screening; gene therapy; antiinflammatory;
KW antiasthmatic; antiarthritic; haemostatic; antiarteriosclerotic;
KW cytostatic; osteopathic; vasotropic; cardiant; virucide; antibacterial;
KW antifungal; vulnerary; antiulcer.
XX
OS Homo sapiens.
XX
PN WO200157188-A2.
XX
PD 09-AUG-2001.
XX
PF 05-FEB-2001; 2001WO-US003800.
XX
PR 03-FEB-2000; 2000US-00496914.
PR 27-APR-2000; 2000US-00560875.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Drmanac RT;
XX
DR WPI; 2001-457740/49.
DR N-PSDB; ABA09052.
XX
PT Human proteins and DNA encoding sequences useful for preventing, treating
PT or ameliorating a medical condition in a mammalian subject e.g. arthritis
PT and cancer.
XX
PS Claim 20; Page 256-257; 1963pp; English.
XX
CC Sequences ABB10981-ABB12330 represent 1350 novel human polypeptides, and
CC sequences ABA08225-ABA09574 represent nucleic acids encoding them. The
CC invention also relates to vectors and recombinant host cells comprising a
CC nucleotide of the invention, methods of producing the novel polypeptides,
CC antibodies against the polypeptides, methods of detecting the nucleotides
CC or polypeptides in a sample, and methods of identifying compounds which
CC bind to polypeptides of the invention. Although novel, many of the
CC polypeptides of the invention have homology to known proteins, thereby
CC giving an insight into their probable biological activities, and hence
CC potential therapeutic applications. The polypeptides of the invention may
CC have various activities, including cytokine, cell proliferation or cell
CC differentiation activities; stem cell growth factor activity;
CC haematopoiesis regulatory activity; tissue growth activity;
CC immunomodulatory activity; activin- or inhibin-related activities;
CC chemotactic or chemokinetic activities; haemostatic, thrombotic or
CC thrombolytic activities; receptor or ligand activities; or may be
CC involved in oncogenesis, cancer cell proliferation or metastasis.
CC Depending on their biological activities, polypeptides and nucleotides of
CC the invention are useful for preventing, treating or ameliorating medical
CC conditions, e.g., by protein or gene therapy. Such conditions include
CC cancers, haematopoietic disorders (e.g.,myeloid or lymphoid cell
CC disorders), chronic inflammatory conditions (e.g., asthma or arthritis),
CC proliferative retinopathy, atherosclerosis, coronary heart disease,
CC arterial ischaemia, bone disorders (e.g., osteoporosis), and abnormal
CC vascular growth. Polypeptides involved with tissue regeneration and
CC repair (or nucleic acids encoding them) may be used to promote wound
CC healing (e.g., of burns, incisions and ulcers), while those with

```
CC   immunomodulatory activities may be used in the treatment of viral,
CC   bacterial and fungal infections in addition to immune disorders.
CC   Polypeptides with growth factor activity may be used in cell cultures to
CC   promote cell growth. For example, such polypeptides may be used to
CC   manipulate stem cells in culture to give rise to neuroepithelial cells
CC   that can be used to augment or replace cells damaged by illness,
CC   autoimmune disease or accidental damage. The polypeptides and nucleotides
CC   may also be used in the diagnosis of the above conditions, and in drug
CC   screening techniques. The present sequence represents a novel human
CC   polypeptide of the invention
XX
SQ   Sequence 1435 AA;

  Query Match             100.0%;  Score 136;  DB 4;  Length 1435;
  Best Local Similarity   100.0%;  Pred. No. 5.7e-12;
  Matches   28;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 VYVFSTEMANKAAEAVLKGQVETIVSFH 28
              ||||||||||||||||||||||||||||
Db        217 VYVFSTEMANKAAEAVLKGQVETIVSFH 244


RESULT 8
AAO05855
ID   AAO05855 standard; protein; 140 AA.
XX
AC   AAO05855;
XX
DT   06-NOV-2001  (first entry)
XX
DE   Human polypeptide SEQ ID NO 19747.
XX
KW   Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW   vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW   tissue growth factor; immunomodulatory; cancer; leukaemia;
KW   nervous system disorders; arthritis; inflammation.
XX
OS   Homo sapiens.
XX
PN   WO200164835-A2.
XX
PD   07-SEP-2001.
XX
PF   26-FEB-2001; 2001WO-US004927.
XX
PR   28-FEB-2000; 2000US-00515126.
PR   18-MAY-2000; 2000US-00577409.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Tang YT,  Liu C,  Drmanac RT;
XX
DR   WPI; 2001-514838/56.
DR   N-PSDB; AAI85786.
XX
PT   Isolated nucleic acids and polypeptides, useful for preventing diagnosing
PT   and treating e.g. leukemia, inflammation and immune disorders.
XX
PS   Claim 20; SEQ ID NO 19747; 1399pp + Sequence Listing; English.
XX
CC   The invention relates to human polynucleotides (AAI79941-AAI93841) and
CC   the encoded proteins (AAO00010-AAO13910) that exhibit activity elating to
CC   cytokine, cell proliferation or cell differentiation or which may induce
CC   production of other cytokines in other cell populations. The
CC   polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC   peptide therapy. The polypeptides have various cytokine-like activities,
CC   e.g. stem cell growth factor activity, haematopoiesis regulating
CC   activity, tissue growth factor activity, immunomodulatory activity and
CC   activin/inhibin activity and may be useful in the diagnosis and/or
CC   treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC   inflammation. Note: The sequence data for this patent did not form part
CC   of the printed specification, but was obtained in electronic format
CC   directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 140 AA;

  Query Match             94.9%;  Score 129;  DB 4;  Length 140;
  Best Local Similarity   92.9%;  Pred. No. 4.5e-12;
  Matches   26;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 VYVFSTEMANKAAEAVLKGQVETIVSFH 28
              |||||||||||||:||||||||:||||
Db         59 VYVFSTEMANKAAKAVLKGQVETMVSFH 86


RESULT 9
ABW01528
ID   ABW01528 standard; peptide; 28 AA.
XX
AC   ABW01528;
XX
DT   15-JAN-2004  (first entry)
XX
DE   Drosophila species legless (lgs) peptide #1.
XX
KW   Legless protein; lgs; cell fate disorder; blood disease; gene therapy;
KW   cancer; tissue regeneration; tissue repair; cytostatic.
XX
OS   Drosophila sp.
XX
PN   US2003114413-A1.
XX
PD   19-JUN-2003.
XX
PF   19-DEC-2002; 2002US-00322579.
XX
PR   28-JUL-2000; 2000US-0221502P.
PR   27-JUL-2001; 2001US-00915543.
XX
PA   (UYZU-) UNIV ZURICH.
XX
PI   Basler K,  Brunner E,  Froesch B,  Kramps T,  Peter O;
XX
DR   WPI; 2003-829432/77.
XX
PT   Novel Lgs polypeptide useful for isolation of Lgs-binding proteins,
PT   diagnosing disorders of cell fate, treating diseases such as cancer.
XX
PS   Claim 28; Fig 7B; 0pp; English.
XX
CC   The invention relates to novel legless (lgs) proteins and polynucleotides
CC   encoding such proteins. Lgs sequences are useful for the treatment of
CC   disorders of cell fate such as differentiation or proliferation. The
CC   invention is used to treat blood disease or a cancerous condition
CC   characterised by over-stimulation of the Wnt pathway such as colon,
CC   breast, head and neck, brain, thyroid, medulloblastoma or skin cancer and
CC   is administered to prevent progression from a pre-neoplastic or non-
CC   malignant condition to a neoplastic or malignant state. It is
CC   administered to promote tissue regeneration and repair. The invention is
CC   also useful in the therapy of diseases cost by an over-activation of Wg
CC   pathway. It is useful for reducing lgs gene expression in an invertebrate
CC   or vertebrate organism or an invertebrate or vertebrate cell line. The
CC   invention is also useful in gene therapy. The present sequence is
CC   Drosophila species legless (lgs) peptide
XX
SQ   Sequence 28 AA;

  Query Match             72.1%;  Score 98;  DB 7;  Length 28;
  Best Local Similarity   57.1%;  Pred. No. 4.9e-08;
  Matches   16;  Conservative    9;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          1 VYVFSTEMANKAAEAVLKGQVETIVSFH 28
              ::||||::|||  ||:|| || :||:::|
Db          1 IFVFSTQLANKGAESVLSGQFQTIIAYH 28
```

```
RESULT 10
ADJ71890
ID   ADJ71890 standard; peptide; 28 AA.
XX
AC   ADJ71890;
XX
DT   20-MAY-2004  (first entry)
XX
DE   Fruit fly legless (lgs) peptide fragment #1.
XX
KW   Fruit fly; legless; lgs; cell differentiation disorder;
KW   cell proliferation disorder; cancer; Wnt pathway; medulloblastoma; colon;
KW   breast; head; neck; brain; thyroid; skin; blood disease;
KW   tissue regeneration; tissue repair; cytostatic.
XX
OS   Drosophila melanogaster.
XX
PN   US2004038901-A1.
XX
PD   26-FEB-2004.
XX
PF   22-SEP-2003; 2003US-00664859.
XX
PR   28-JUL-2000; 2000US-0221502P.
PR   27-JUL-2001; 2001US-00915543.
XX
PA   (UYZU-) UNIV ZURICH.
XX
PI   Basler K,  Brunner E,  Froesch B,  Kramps T,  Peter O;
XX
DR   WPI; 2004-203288/19.
XX
PT   Novel polypeptide sharing one or more homologue amino acid domains with
PT   Legless protein being functional homologue of Legless, useful for
PT   diagnosing disorders of cell fate.
XX
PS   Disclosure; SEQ ID NO 2; 62pp; English.
XX
CC   The invention relates to a polypeptide sharing one or more homologous
CC   amino acid domains with a Legless (Lgs) protein and is therefore a
CC   functional homologue of Lgs. The invention also relates to a nucleotide
CC   sequence encoding a protein present in invertebrate and/or vertebrate
CC   organisms, the nucleotide sequence encoding a protein comprising a
CC   positive function in a regulatory pathway and the use of the polypeptide
CC   for the isolation of Lgs-binding proteins by carrying out an assay chosen
CC   from an in vitro binding assay with such a peptide or a co-
CC   immunoprecipitation from vertebrate or invertebrate cell lysates or a
CC   mammalian or yeast two hybrid assay. The polypeptide and polynucleotide
CC   are useful for treating disorders of cell fate, which involves
CC   administering therapeutic compounds chosen from invertebrate and
CC   vertebrate Lgs protein homologues or fragments, antibodies, antibody
CC   fragments, lgs antisense DNA, lgs antisense RNA, lgs double-stranded RNA,
CC   small peptides or chemical and natural compounds being capable of
CC   interfering with Lgs function, synthesis and degradation. The disorders
CC   are related to cell differentiation or cell proliferation. The compound
CC   is administered to treat a cancerous condition by preventing progression
CC   from a pre-neoplastic or non-malignant condition to a neoplastic or
CC   malignant state. The cancerous condition is characterised by over-
CC   stimulation of the Wnt pathway and is medulloblastoma or cancer of the
CC   colon, breast, head and neck, brain, thyroid or skin. The therapeutic
CC   compound may also be administered to a blood disease to promote tissue
CC   regeneration and repair. This sequence represents a Drosophila Lgs
CC   peptide fragment of the invention.
XX
SQ   Sequence 28 AA;

  Query Match             72.1%;  Score 98;  DB 8;  Length 28;
  Best Local Similarity   57.1%;  Pred. No. 4.9e-08;
  Matches   16;  Conservative    9;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          1 VYVFSTEMANKAAEAVLKGQVETIVSFH 28
              ::||||::||| ||:|| || :||:::|
Db          1 IFVFSTQLANKGAESVLSGQFQTIIAYH 28
```

```
RESULT 11
ABB58779
ID   ABB58779 standard; protein; 1429 AA.
XX
AC   ABB58779;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Drosophila melanogaster polypeptide SEQ ID NO 3129.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US009231.
XX
PR   23-MAR-2000; 2000US-0191637P.
PR   11-JUL-2000; 2000US-00614150.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Venter JC,  Adams M,  Li PWD,  Myers EW;
XX
DR   WPI; 2001-656860/75.
DR   N-PSDB; ABL02882.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signaling and cell-cell
PT   interactions.
XX
PS   Disclosure; SEQ ID NO 3129; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-
CC   ABB72072). The sequence data for this patent did not form part of the
CC   printed specification, but was obtained in electronic format directly
CC   from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 1429 AA;

  Query Match             72.1%;  Score 98;  DB 4;  Length 1429;
  Best Local Similarity   57.1%;  Pred. No. 5.4e-06;
  Matches   16;  Conservative    9;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          1 VYVFSTEMANKAAEAVLKGQVETIVSFH 28
              ::||||::||| ||:|| || :||:::|
Db        323 IFVFSTQLANKGAESVLSGQFQTIIAYH 350
```

```
RESULT 12
AAB71228
ID   AAB71228 standard; protein; 1464 AA.
XX
AC   AAB71228;
XX
DT   18-NOV-2002  (first entry)
XX
DE   D. melanogaster lgs protein.
XX
KW   Legless; fruitfly; lgs; Wnt/Wingless signaling pathway; Wnt; Wg;
KW   tissue proliferation; tumour; cytostatic; cellular disorder; colon;
```

KW blood disorder; cancer; breast; head and neck cancer; brain; thyroid;
KW medulloblastoma; skin cancer; tissue regeneration; tissue repair.
XX
OS Drosophila melanogaster.
XX
PN US2002086986-A1.
XX
PD 04-JUL-2002.
XX
PF 27-JUL-2001; 2001US-00915543.
XX
PR 28-JUL-2000; 2000US-0221502P.
XX
PA (BASL/) BASLER K.
PA (BRUN/) BRUNNER E.
PA (FROE/) FROESCH B.
PA (KRAM/) KRAMPS T.
PA (PETE/) PETER O.
XX
PI Basler K, Brunner E, Froesch B, Kramps T, Peter O;
XX
DR WPI; 2002-635689/68.
DR N-PSDB; AAF88466.
XX
PT Novel polypeptide useful in therapeutic method for treating disorders of
PT cell fate such as cell differentiation or cell proliferation.
XX
PS Example II; Fig 2; 41pp; English.
XX
CC This invention describes a novel polypeptide sharing one or more
CC homologous amino acid domains with the legless (lgs) protein, a
CC downstream component of the Wnt/Wingless (Wnt/Wg) signaling pathway
CC involved in the formation and maintenance of spatial arrangements and
CC proliferation of tissues during development, and in the formation and
CC growth of many human tumours. The products of the invention have
CC cytostatic activity and can be used to treat cellular disorders, blood
CC disorders and cancers caused by over-stimulation of the Wnt pathway,
CC where the cancerous condition is colon, breast, head and neck, brain,
CC thyroid, medulloblastoma or skin cancer. The product could also be used
CC to promote tissue regeneration and repair. This sequence represents the
CC Drosophila melanogaster (fruitfly) legless (lgs) protein described in the
CC disclosure of the invention
XX
SQ Sequence 1464 AA;

Query Match 72.1%; Score 98; DB 5; Length 1464;
Best Local Similarity 57.1%; Pred. No. 5.5e-06;
Matches 16; Conservative 9; Mismatches 3; Indels 0; Gaps 0;

Qy 1 VYVFSTEMANKAAEAVLKGQVETIVSFH 28
::||||::||| ||:|| || :||:::|
Db 318 IFVFSTQLANKGAESVLSGQFQTIIAYH 345

RESULT 13
ABW01527
ID ABW01527 standard; protein; 1464 AA.
XX
AC ABW01527;
XX
DT 15-JAN-2004 (first entry)
XX
DE Drosophila species legless (lgs) protein.
XX
KW Legless protein; lgs; cell fate disorder; blood disease; gene therapy;
KW cancer; tissue regeneration; tissue repair; cytostatic.
XX
OS Drosophila sp.
XX
PN US2003114413-A1.
XX
PD 19-JUN-2003.
XX
PF 19-DEC-2002; 2002US-00322579.
XX
PR 28-JUL-2000; 2000US-0221502P.
PR 27-JUL-2001; 2001US-00915543.
XX
PA (UYZU-) UNIV ZURICH.
XX
PI Basler K, Brunner E, Froesch B, Kramps T, Peter O;
XX
DR WPI; 2003-829432/77.
DR N-PSDB; AAD62641.
XX
PT Novel Lgs polypeptide useful for isolation of Lgs-binding proteins,
PT diagnosing disorders of cell fate, treating diseases such as cancer.
XX
PS Claim 5; Fig 2; 0pp; English.
XX
CC The invention relates to novel legless (lgs) proteins and polynucleotides
CC encoding such proteins. Lgs sequences are useful for the treatment of
CC disorders of cell fate such as differentiation or proliferation. The
CC invention is used to treat blood disease or a cancerous condition
CC characterised by over-stimulation of the Wnt pathway such as colon,
CC breast, head and neck, brain, thyroid, medulloblastoma or skin cancer and
CC is administered to prevent progression from a pre-neoplastic or non-
CC malignant condition to a neoplastic or malignant state. It is
CC administered to promote tissue regeneration and repair. The invention is
CC also useful in the therapy of diseases cost by an over-activation of Wg
CC pathway. It is useful for reducing lgs gene expression in an invertebrate
CC or vertebrate organism or an invertebrate or vertebrate cell line. The
CC invention is also useful in gene therapy. The present sequence is
CC Drosophila species legless (lgs) protein
XX
SQ Sequence 1464 AA;

Query Match 72.1%; Score 98; DB 7; Length 1464;
Best Local Similarity 57.1%; Pred. No. 5.5e-06;
Matches 16; Conservative 9; Mismatches 3; Indels 0; Gaps 0;

Qy 1 VYVFSTEMANKAAEAVLKGQVETIVSFH 28
::||||::||| ||:|| || :||:::|
Db 318 IFVFSTQLANKGAESVLSGQFQTIIAYH 345

RESULT 14
ADJ71911
ID ADJ71911 standard; protein; 1464 AA.
XX
AC ADJ71911;
XX
DT 20-MAY-2004 (first entry)
XX
DE Fruit fly legless (lgs) polypeptide.
XX
KW Fruit fly; legless; lgs; cell differentiation disorder;
KW cell proliferation disorder; cancer; Wnt pathway; medulloblastoma; colon;
KW breast; head; neck; brain; thyroid; skin; blood disease;
KW tissue regeneration; tissue repair; cytostatic.
XX
OS Drosophila melanogaster.
XX
PN US2004038901-A1.
XX
PD 26-FEB-2004.
XX
PF 22-SEP-2003; 2003US-00664859.
XX
PR 28-JUL-2000; 2000US-0221502P.
PR 27-JUL-2001; 2001US-00915543.
XX
PA (UYZU-) UNIV ZURICH.
XX
PI Basler K, Brunner E, Froesch B, Kramps T, Peter O;
XX

DR WPI; 2004-203288/19.
DR N-PSDB; ADJ71889.
XX
PT Novel polypeptide sharing one or more homologue amino acid domains with
PT Legless protein being functional homologue of Legless, useful for
PT diagnosing disorders of cell fate.
XX
PS Example 2; SEQ ID NO 1; 62pp; English.
XX
CC The invention relates to a polypeptide sharing one or more homologous
CC amino acid domains with a Legless (Lgs) protein and is therefore a
CC functional homologue of Lgs. The invention also relates to a nucleotide
CC sequence encoding a protein present in invertebrate and/or vertebrate
CC organisms, the nucleotide sequence encoding a protein comprising a
CC positive function in a regulatory pathway and the use of the polypeptide
CC for the isolation of Lgs-binding proteins by carrying out an assay chosen
CC from an in vitro binding assay with such a peptide or a co-
CC immunoprecipitation from vertebrate or invertebrate cell lysates or a
CC mammalian or yeast two hybrid assay. The polypeptide and polynucleotide
CC are useful for treating disorders of cell fate, which involves
CC administering therapeutic compounds chosen from invertebrate and
CC vertebrate Lgs protein homologues or fragments, antibodies, antibody
CC fragments, lgs antisense DNA, lgs antisense RNA, lgs double-stranded RNA,
CC small peptides or chemical and natural compounds being capable of
CC interfering with Lgs function, synthesis and degradation. The disorders
CC are related to cell differentiation or cell proliferation. The compound
CC is administered to treat a cancerous condition by preventing progression
CC from a pre-neoplastic or non-malignant condition to a neoplastic or
CC malignant state. The cancerous condition is characterised by over-
CC stimulation of the Wnt pathway and is medulloblastoma or cancer of the
CC colon, breast, head and neck, brain, thyroid or skin. The therapeutic
CC compound may also be administered to a blood disease to promote tissue
CC regeneration and repair. This sequence represents the Drosophila Lgs
CC polypeptide of the invention.
XX
SQ Sequence 1464 AA;

Query Match 72.1%; Score 98; DB 8; Length 1464;
Best Local Similarity 57.1%; Pred. No. 5.5e-06;
Matches 16; Conservative 9; Mismatches 3; Indels 0; Gaps 0;

Qy 1 VYVFSTEMANKAAEAVLKGQVETIVSFH 28
::||||::||| ||:|| || :||:::|
Db 318 IFVFSTQLANKGAESVLSGQFQTIIAYH 345

RESULT 15
AAU78460
ID AAU78460 standard; protein; 1494 AA.
XX
AC AAU78460;
XX
DT 02-JUL-2002 (first entry)
XX
DE Mouse beta-catenin nuclear localised protein.
XX
KW Mouse; beta-catenin nuclear localised protein; cancer; gene therapy; EST;
KW expressed sequence tag.
XX
OS Mus musculus.
XX
PN WO200224738-A1.
XX
PD 28-MAR-2002.
XX
PF 19-SEP-2001; 2001WO-JP008140.
XX
PR 22-SEP-2000; 2000JP-00287876.
XX
PA (KYOW) KYOWA HAKKO KOGYO KK.
XX
PI Akiyama T, Adachi S;
XX
DR WPI; 2002-330014/36.
DR N-PSDB; ABK47631.
XX
PT New beta-catenin nuclear localized protein for diagnosis and treatment of
PT diseases associated with nuclear localization of beta-catenin e.g.
PT cancer.
XX
PS Claim 1; Page 81-88; 113pp; Japanese.
XX
CC The invention relates to a beta-catenin nuclear localised protein and DNA
CC encoding the protein. The protein and encoding DNA are applicable in
CC diagnosis and treatment of diseases associated with nuclear localisation
CC of beta-catenin e.g. cancer, including gene therapy. The present sequence
CC represents the amino acid sequence of mouse beta-catenin nuclear
CC localised protein
XX
SQ Sequence 1494 AA;

Query Match 69.1%; Score 94; DB 5; Length 1494;
Best Local Similarity 60.7%; Pred. No. 2.4e-05;
Matches 17; Conservative 8; Mismatches 3; Indels 0; Gaps 0;

Qy 1 VYVFSTEMANKAAEAVLKGQVETIVSFH 28
||||:| :|| ||||||:|: |:|:::|
Db 238 VYVFTTHLANTAAEAVLQGRAESILAYH 265

RESULT 16
ABP06595
ID ABP06595 standard; protein; 114 AA.
XX
AC ABP06595;
XX
DT 25-JUN-2002 (first entry)
XX
DE Human ORFX protein sequence SEQ ID NO:13172.
XX
KW Human; open reading frame; ORFX; gene therapy; cancer; cirrhosis;
KW hyperproliferative disorder; psoriasis; benign tumour; haemorrhage;
KW degenerative disorder; osteoarthritis; neurodegenerative disorder;
KW cardiovascular disease; diabetes mellitus; systemic lupus erythematosus;
KW hypertension; hypothyroidism; cholesterol ester storage disease;
KW immune deficiency; immune disorder; infectious disease;
KW autoimmune disorder; rheumatoid arthritis; autoimmune thyroiditis;
KW myasthenia gravis.
XX
OS Homo sapiens.
XX
PN WO200192523-A2.
XX
PD 06-DEC-2001.
XX
PF 29-MAY-2001; 2001WO-US010836.
XX
PR 30-MAY-2000; 2000US-0206132P.
PR 29-AUG-2000; 2000US-0228716P.
XX
PA (CURA-) CURAGEN CORP.
XX
PI Shimkets RA, Leach MD;
XX
DR WPI; 2002-106308/14.
DR N-PSDB; ABN22347.
XX
PT Novel human polypeptides and polynucleotides useful for diagnosing,
PT preventing and treating cardiovascular disease, neurodegenerative,
PT hyperproliferative disorders and autoimmune disorders.
XX
PS Disclosure; SEQ ID NO 13172; 1037pp; English.
XX
CC The present invention describes substantially purified human proteins
CC (referred to as open reading frame, ORFX, where X is 1-11491 (see Table 1
CC in the specification). ABN15762 to ABN27252 encode the human ORFX

```
CC   proteins given in ABP00010 to ABP11500. ORFX proteins are useful for
CC   treating or preventing a pathology associated with an ORFX-associated
CC   disorder in humans, and in the manufacture of a medicament for treating a
CC   syndrome associated with ORFX-associated disorder. ORFX polynucleotide
CC   sequences can be used in gene therapy. ORFX sequences can be used in the
CC   treatment of cancer, hyperproliferative disorders, cirrhosis of liver,
CC   psoriasis, benign tumours, keloid, degenerative disorders, haemorrhage,
CC   osteoarthritis, neurodegenerative disorders, disorders related to organ
CC   transplantation, cardiovascular diseases, diabetes mellitus, systemic
CC   lupus erythematosus, hypertension, hypothyroidism, cholesterol ester
CC   storage disease, various immune deficiencies and disorders, infectious
CC   diseases, autoimmune disorders such as multiple sclerosis, rheumatoid
CC   arthritis, autoimmune thyroiditis, myasthenia gravis, graft-versus-host
CC   disease and autoimmune inflammatory eye disease. ORFX proteins are also
CC   useful for treating burns, incisions, ulcers, for treating osteoporosis,
CC   bone degenerative disorders, or periodontal disease, and for gut
CC   protection or regeneration and treatment of lung or liver fibrosis,
CC   reperfusion injury in various tissues and conditions resulting from
CC   systemic cytokine damage. N.B. The sequence data for this patent did not
CC   form part of the printed specification, but was obtained in electronic
CC   format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 114 AA;

  Query Match             66.9%;  Score 91;  DB 5;  Length 114;
  Best Local Similarity   57.1%;  Pred. No. 3.3e-06;
  Matches   16;  Conservative    9;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          1 VYVFSTEMANKAAEAVLKGQVETIVSFH 28
              ||||:| :|| ||||||:|: ::|:::|
Db         78 VYVFTTHLANTAAEAVLQGRADSILAYH 105


RESULT 17
AAU78461
ID   AAU78461 standard; protein; 320 AA.
XX
AC   AAU78461;
XX
DT   02-JUL-2002  (first entry)
XX
DE   Mouse beta-catenin nuclear localised protein #2.
XX
KW   Mouse; beta-catenin nuclear localised protein; cancer; gene therapy; EST;
KW   expressed sequence tag.
XX
OS   Mus musculus.
XX
PN   WO200224738-A1.
XX
PD   28-MAR-2002.
XX
PF   19-SEP-2001; 2001WO-JP008140.
XX
PR   22-SEP-2000; 2000JP-00287876.
XX
PA   (KYOW ) KYOWA HAKKO KOGYO KK.
XX
PI   Akiyama T,  Adachi S;
XX
DR   WPI; 2002-330014/36.
DR   N-PSDB; ABK47632.
XX
PT   New beta-catenin nuclear localized protein for diagnosis and treatment of
PT   diseases associated with nuclear localization of beta-catenin e.g.
PT   cancer.
XX
PS   Claim 2; Page 91-92; 113pp; Japanese.
XX
CC   The invention relates to a beta-catenin nuclear localised protein and DNA
CC   encoding the protein. The protein and encoding DNA are applicable in
CC   diagnosis and treatment of diseases associated with nuclear localisation
CC   of beta-catenin e.g. cancer, including gene therapy. The present sequence
CC   represents the amino acid sequence of mouse beta-catenin nuclear
CC   localised protein #2
XX
SQ   Sequence 320 AA;

  Query Match             49.3%;  Score 67;  DB 5;  Length 320;
  Best Local Similarity   57.1%;  Pred. No. 0.068;
  Matches   12;  Conservative    7;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          8 MANKAAEAVLKGQVETIVSFH 28
              :|| ||||||:|: |:|:::|
Db          1 LANTAAEAVLQGRAESILAYH 21


RESULT 18
AAO07544
ID   AAO07544 standard; protein; 113 AA.
XX
AC   AAO07544;
XX
DT   06-NOV-2001  (first entry)
XX
DE   Human polypeptide SEQ ID NO 21436.
XX
KW   Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW   vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW   tissue growth factor; immunomodulatory; cancer; leukaemia;
KW   nervous system disorders; arthritis; inflammation.
XX
OS   Homo sapiens.
XX
PN   WO200164835-A2.
XX
PD   07-SEP-2001.
XX
PF   26-FEB-2001; 2001WO-US004927.
XX
PR   28-FEB-2000; 2000US-00515126.
PR   18-MAY-2000; 2000US-00577409.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Tang YT,  Liu C,  Drmanac RT;
XX
DR   WPI; 2001-514838/56.
DR   N-PSDB; AAI87475.
XX
PT   Isolated nucleic acids and polypeptides, useful for preventing diagnosing
PT   and treating e.g. leukemia, inflammation and immune disorders.
XX
PS   Claim 20; SEQ ID NO 21436; 1399pp + Sequence Listing; English.
XX
CC   The invention relates to human polynucleotides (AAI79941-AAI93841) and
CC   the encoded proteins (AAO00010-AAO13910) that exhibit activity elating to
CC   cytokine, cell proliferation or cell differentiation or which may induce
CC   production of other cytokines in other cell populations. The
CC   polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC   peptide therapy. The polypeptides have various cytokine-like activities,
CC   e.g. stem cell growth factor activity, haematopoiesis regulating
CC   activity, tissue growth factor activity, immunomodulatory activity and
CC   activin/inhibin activity and may be useful in the diagnosis and/or
CC   treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC   inflammation. Note: The sequence data for this patent did not form part
CC   of the printed specification, but was obtained in electronic format
CC   directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 113 AA;

  Query Match             47.8%;  Score 65;  DB 4;  Length 113;
  Best Local Similarity   75.0%;  Pred. No. 0.041;
  Matches   15;  Conservative    1;  Mismatches    4;  Indels    0;  Gaps    0;

Qy          8 MANKAAEAVLKGQVETIVSF 27
```

```
                |  :||||||||||||  ||
Db         81 MCVQAAEAVLKGQVETDASF 100


RESULT 19
ADN48110
ID   ADN48110 standard; protein; 133 AA.
XX
AC   ADN48110;
XX
DT   01-JUL-2004  (first entry)
XX
DE   Thermococcus kodakaraensis KOD1 protein sequence SeqID1988.
XX
KW   gene disruption; gene targeting; marker gene; transformation;
KW   homologous recombination; hyperthermostable archaebacterium; KOD1;
KW   gene structure; gene function; enzyme activity; medicine;
KW   forensic science; food; drug inspection; molecular biology; immunology.
XX
OS   Thermococcus kodakaraensis.
XX
PN   WO2004022736-A1.
XX
PD   18-MAR-2004.
XX
PF   29-AUG-2003; 2003WO-IB003597.
XX
PR   30-AUG-2002; 2002JP-00319011.
XX
PA   (NISC-) JAPAN SCI & TECHNOLOGY CORP.
XX
PI   Imanaka T,  Atomi H;
XX
DR   WPI; 2004-257583/24.
XX
PT   Method for disrupting targeted gene in genome of organism particularly
PT   thermostable bacterium and with genome chips for analysis, applicable in
PT   studying gene structure and functions.
XX
PS   Claim 9; SEQ ID NO 1988; 598pp; Japanese.
XX
CC   This invention relates to a novel method for targeting disruption of an
CC   arbitrary gene in a genome of an organism which comprises providing the
CC   whole sequential data of the genome of such organism, selecting at least
CC   1 arbitrary region in the sequence, providing a vector that contains a
CC   sequence homologous with the selected region and a marker gene,
CC   transformation, and homologous recombination. The genome is preferably
CC   the genome of a hyperthermostable archaebacterium, particularly
CC   Thermococcus kodakaraensis KOD1. The method is for targeting the
CC   disruption of a gene in the genome of an organism, which is applicable in
CC   studying gene structure and functions as well as enzyme activities of
CC   encoded proteins and useful in medicine, forensic science, food or drug
CC   inspection, molecular biology and immunology. With this method, the
CC   disruption of a gene at an arbitrary position in a genome can be achieved
CC   efficiently and reliably. The present sequence is that of a protein
CC   encoded by the genome of Thermococcus kodakaraensis which was derived
CC   using the method of the invention. Note: The sequence data for this
CC   patent did not form part of the printed specification, but was obtained
CC   in electronic format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 133 AA;

  Query Match             39.0%;  Score 53;  DB 8;  Length 133;
  Best Local Similarity   37.5%;  Pred. No. 3.8;
  Matches    9;  Conservative    9;  Mismatches    6;  Indels    0;  Gaps    0;

Qy          3 VFSTEMANKAAEAVLKGQVETIVS 26
              ::: :  ||| :|:: |: ||||:
Db         69 IYALKKNNKAPKAIIVGEAETIVA 92


RESULT 20
ABO66692
ID   ABO66692 standard; protein; 322 AA.
XX
AC   ABO66692;
XX
DT   29-JUL-2004  (first entry)
XX
DE   Klebsiella pneumoniae polypeptide seqid 13209.
XX
KW   Recombinant expression vector; transcription regulatory element;
KW   Klebsiella pneumoniae protein; antibacterial; Vaccine.
XX
OS   Klebsiella pneumoniae.
XX
PN   US6610836-B1.
XX
PD   26-AUG-2003.
XX
PF   27-JAN-2000; 2000US-00489039.
XX
PR   29-JAN-1999;   99US-0117747P.
XX
PA   (GENO-) GENOME THERAPEUTICS CORP.
XX
PI   Breton GL,  Osborne M;
XX
DR   WPI; 2003-895346/82.
DR   N-PSDB; ABD00263.
XX
PT   New nucleic acid encoding a Klebsiella pneumoniae polypeptide, useful for
PT   preparing a vaccine composition against Klebsiella pneumoniae.
XX
PS   Disclosure; SEQ ID NO 13209; 932pp; English.
XX
CC   The invention describes a new isolated nucleic acid encoding a Klebsiella
CC   pneumoniae polypeptide. Also described are: a recombinant expression
CC   vector comprising the nucleic acid, operably linked to a transcription
CC   regulatory element; and a cell comprising the recombinant expression
CC   vector. The nucleic acid is useful for preparing a vaccine composition
CC   against Klebsiella pneumoniae. This is the amino acid sequence of a
CC   Klebsiella pneumoniae polypeptide of the invention
XX
SQ   Sequence 322 AA;

  Query Match             39.0%;  Score 53;  DB 7;  Length 322;
  Best Local Similarity   42.3%;  Pred. No. 11;
  Matches   11;  Conservative    7;  Mismatches    8;  Indels    0;  Gaps    0;

Qy          3 VFSTEMANKAAEAVLKGQVETIVSFH 28
              ::  ::| :|||||| |: :|  | |
Db        128 LYRQDLAPEAAEAVLHGKWDTQASMH 153


RESULT 21
AAG33446
ID   AAG33446 standard; protein; 360 AA.
XX
AC   AAG33446;
XX
DT   18-OCT-2000  (first entry)
XX
DE   Zea mays protein fragment SEQ ID NO: 40525.
XX
KW   Protein identification; signal transduction pathway; metabolic pathway;
KW   hybridisation assay; genetic mapping; gene expression control; promoter;
KW   termination sequence; corn.
XX
OS   Zea mays subsp. mays.
XX
PN   EP1033405-A2.
XX
PD   06-SEP-2000.
XX
```

PF 25-FEB-2000; 2000EP-00301439.
XX
PR 25-FEB-1999; 99US-0121825P.
PR 05-MAR-1999; 99US-0123180P.
PR 09-MAR-1999; 99US-0123548P.
PR 23-MAR-1999; 99US-0125788P.
PR 25-MAR-1999; 99US-0126264P.
PR 29-MAR-1999; 99US-0126785P.
PR 01-APR-1999; 99US-0127462P.
PR 06-APR-1999; 99US-0128234P.
PR 08-APR-1999; 99US-0128714P.
PR 16-APR-1999; 99US-0129845P.
PR 19-APR-1999; 99US-0130077P.
PR 21-APR-1999; 99US-0130449P.
PR 23-APR-1999; 99US-0130510P.
PR 23-APR-1999; 99US-0130891P.
PR 28-APR-1999; 99US-0131449P.
PR 30-APR-1999; 99US-0132048P.
PR 30-APR-1999; 99US-0132407P.
PR 04-MAY-1999; 99US-0132484P.
PR 05-MAY-1999; 99US-0132485P.
PR 06-MAY-1999; 99US-0132486P.
PR 06-MAY-1999; 99US-0132487P.
PR 07-MAY-1999; 99US-0132863P.
PR 11-MAY-1999; 99US-0134256P.
PR 14-MAY-1999; 99US-0134218P.
PR 14-MAY-1999; 99US-0134219P.
PR 14-MAY-1999; 99US-0134221P.
PR 14-MAY-1999; 99US-0134370P.
PR 18-MAY-1999; 99US-0134768P.
PR 19-MAY-1999; 99US-0134941P.
PR 20-MAY-1999; 99US-0135124P.
PR 21-MAY-1999; 99US-0135353P.
PR 24-MAY-1999; 99US-0135629P.
PR 25-MAY-1999; 99US-0136021P.
PR 27-MAY-1999; 99US-0136392P.
PR 28-MAY-1999; 99US-0136782P.
PR 01-JUN-1999; 99US-0137222P.
PR 03-JUN-1999; 99US-0137528P.
PR 04-JUN-1999; 99US-0137502P.
PR 07-JUN-1999; 99US-0137724P.
PR 08-JUN-1999; 99US-0138094P.
PR 10-JUN-1999; 99US-0138540P.
PR 10-JUN-1999; 99US-0138847P.
PR 14-JUN-1999; 99US-0139119P.
PR 16-JUN-1999; 99US-0139452P.
PR 16-JUN-1999; 99US-0139453P.
PR 17-JUN-1999; 99US-0139492P.
PR 18-JUN-1999; 99US-0139454P.
PR 18-JUN-1999; 99US-0139455P.
PR 18-JUN-1999; 99US-0139456P.
PR 18-JUN-1999; 99US-0139457P.
PR 18-JUN-1999; 99US-0139458P.
PR 18-JUN-1999; 99US-0139459P.
PR 18-JUN-1999; 99US-0139460P.
PR 18-JUN-1999; 99US-0139461P.
PR 18-JUN-1999; 99US-0139462P.
PR 18-JUN-1999; 99US-0139463P.
PR 18-JUN-1999; 99US-0139750P.
PR 18-JUN-1999; 99US-0139763P.
PR 21-JUN-1999; 99US-0139817P.
PR 22-JUN-1999; 99US-0139899P.
PR 23-JUN-1999; 99US-0140353P.
PR 23-JUN-1999; 99US-0140354P.
PR 24-JUN-1999; 99US-0140695P.
PR 28-JUN-1999; 99US-0140823P.
PR 29-JUN-1999; 99US-0140991P.
PR 30-JUN-1999; 99US-0141287P.
PR 01-JUL-1999; 99US-0141842P.
PR 01-JUL-1999; 99US-0142154P.
PR 02-JUL-1999; 99US-0142055P.
PR 06-JUL-1999; 99US-0142390P.
PR 08-JUL-1999; 99US-0142803P.
PR 09-JUL-1999; 99US-0142920P.
PR 12-JUL-1999; 99US-0142977P.
PR 13-JUL-1999; 99US-0143542P.
PR 14-JUL-1999; 99US-0143624P.
PR 15-JUL-1999; 99US-0144005P.
PR 16-JUL-1999; 99US-0144085P.
PR 16-JUL-1999; 99US-0144086P.
PR 19-JUL-1999; 99US-0144325P.
PR 19-JUL-1999; 99US-0144331P.
PR 19-JUL-1999; 99US-0144332P.
PR 19-JUL-1999; 99US-0144333P.
PR 19-JUL-1999; 99US-0144334P.
PR 19-JUL-1999; 99US-0144335P.
PR 20-JUL-1999; 99US-0144352P.
PR 20-JUL-1999; 99US-0144632P.
PR 20-JUL-1999; 99US-0144884P.
PR 21-JUL-1999; 99US-0144814P.
PR 21-JUL-1999; 99US-0145086P.
PR 21-JUL-1999; 99US-0145088P.
PR 22-JUL-1999; 99US-0145085P.
PR 22-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145089P.
PR 22-JUL-1999; 99US-0145192P.
PR 23-JUL-1999; 99US-0145145P.
PR 23-JUL-1999; 99US-0145218P.
PR 23-JUL-1999; 99US-0145224P.
PR 26-JUL-1999; 99US-0145276P.
PR 27-JUL-1999; 99US-0145913P.
PR 27-JUL-1999; 99US-0145918P.
PR 27-JUL-1999; 99US-0145919P.
PR 28-JUL-1999; 99US-0145951P.
PR 02-AUG-1999; 99US-0146386P.
PR 02-AUG-1999; 99US-0146388P.
PR 02-AUG-1999; 99US-0146389P.
PR 03-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147204P.
PR 04-AUG-1999; 99US-0147302P.
PR 05-AUG-1999; 99US-0147192P.
PR 05-AUG-1999; 99US-0147260P.
PR 06-AUG-1999; 99US-0147303P.
PR 06-AUG-1999; 99US-0147416P.
PR 09-AUG-1999; 99US-0147493P.
PR 09-AUG-1999; 99US-0147935P.
PR 10-AUG-1999; 99US-0148171P.
PR 11-AUG-1999; 99US-0148319P.
PR 12-AUG-1999; 99US-0148341P.
PR 13-AUG-1999; 99US-0148565P.
PR 13-AUG-1999; 99US-0148684P.
PR 16-AUG-1999; 99US-0149368P.
PR 17-AUG-1999; 99US-0149175P.
PR 18-AUG-1999; 99US-0149426P.
PR 20-AUG-1999; 99US-0149722P.
PR 20-AUG-1999; 99US-0149723P.
PR 20-AUG-1999; 99US-0149929P.
PR 23-AUG-1999; 99US-0149902P.
PR 23-AUG-1999; 99US-0149930P.
PR 25-AUG-1999; 99US-0150566P.
PR 26-AUG-1999; 99US-0150884P.
PR 27-AUG-1999; 99US-0151065P.
PR 27-AUG-1999; 99US-0151066P.
PR 27-AUG-1999; 99US-0151080P.
PR 30-AUG-1999; 99US-0151303P.
PR 31-AUG-1999; 99US-0151438P.
PR 01-SEP-1999; 99US-0151930P.
PR 07-SEP-1999; 99US-0152363P.
PR 10-SEP-1999; 99US-0153070P.
PR 13-SEP-1999; 99US-0153758P.
PR 15-SEP-1999; 99US-0154018P.
PR 16-SEP-1999; 99US-0154039P.
PR 20-SEP-1999; 99US-0154779P.
PR 22-SEP-1999; 99US-0155139P.
PR 23-SEP-1999; 99US-0155486P.
PR 24-SEP-1999; 99US-0155659P.

PR 28-SEP-1999; 99US-0156458P.
PR 29-SEP-1999; 99US-0156596P.
PR 04-OCT-1999; 99US-0157117P.
PR 05-OCT-1999; 99US-0157753P.
PR 06-OCT-1999; 99US-0157865P.
PR 07-OCT-1999; 99US-0158029P.
PR 08-OCT-1999; 99US-0158232P.
PR 12-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0159293P.
PR 13-OCT-1999; 99US-0159294P.
PR 13-OCT-1999; 99US-0159295P.
PR 14-OCT-1999; 99US-0159329P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159637P.
PR 14-OCT-1999; 99US-0159638P.
PR 18-OCT-1999; 99US-0159584P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160768P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 21-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160989P.
PR 25-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161406P.
PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161992P.
PR 28-OCT-1999; 99US-0161993P.
PR 29-OCT-1999; 99US-0162142P.

Query Match 38.2%; Score 52; DB 3; Length 360;
Best Local Similarity 50.0%; Pred. No. 18;
Matches 12; Conservative 2; Mismatches 10; Indels 0; Gaps 0;

```
Qy          4 FSTEMANKAAEAVLKGQVETIVSF 27
              ||| :||     |||   ||| :|
Db         97 FSTRLANNLENLVLKEGPETIAAF 120
```

RESULT 22
AAG33445
ID AAG33445 standard; protein; 448 AA.
XX
AC AAG33445;
XX
DT 18-OCT-2000 (first entry)
XX
DE Zea mays protein fragment SEQ ID NO: 40524.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence; corn.
XX
OS Zea mays subsp. mays.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-00301439.
XX
PR 25-FEB-1999; 99US-0121825P.
PR 05-MAR-1999; 99US-0123180P.
PR 09-MAR-1999; 99US-0123548P.
PR 23-MAR-1999; 99US-0125788P.
PR 25-MAR-1999; 99US-0126264P.
PR 29-MAR-1999; 99US-0126785P.
PR 01-APR-1999; 99US-0127462P.
PR 06-APR-1999; 99US-0128234P.
PR 08-APR-1999; 99US-0128714P.
PR 16-APR-1999; 99US-0129845P.
PR 19-APR-1999; 99US-0130077P.
PR 21-APR-1999; 99US-0130449P.
PR 23-APR-1999; 99US-0130510P.
PR 23-APR-1999; 99US-0130891P.
PR 28-APR-1999; 99US-0131449P.
PR 30-APR-1999; 99US-0132048P.
PR 30-APR-1999; 99US-0132407P.
PR 04-MAY-1999; 99US-0132484P.
PR 05-MAY-1999; 99US-0132485P.
PR 06-MAY-1999; 99US-0132486P.
PR 06-MAY-1999; 99US-0132487P.
PR 07-MAY-1999; 99US-0132863P.
PR 11-MAY-1999; 99US-0134256P.
PR 14-MAY-1999; 99US-0134218P.
PR 14-MAY-1999; 99US-0134219P.
PR 14-MAY-1999; 99US-0134221P.
PR 14-MAY-1999; 99US-0134370P.
PR 18-MAY-1999; 99US-0134768P.
PR 19-MAY-1999; 99US-0134941P.
PR 20-MAY-1999; 99US-0135124P.
PR 21-MAY-1999; 99US-0135353P.
PR 24-MAY-1999; 99US-0135629P.
PR 25-MAY-1999; 99US-0136021P.
PR 27-MAY-1999; 99US-0136392P.
PR 28-MAY-1999; 99US-0136782P.
PR 01-JUN-1999; 99US-0137222P.
PR 03-JUN-1999; 99US-0137528P.
PR 04-JUN-1999; 99US-0137502P.
PR 07-JUN-1999; 99US-0137724P.
PR 08-JUN-1999; 99US-0138094P.
PR 10-JUN-1999; 99US-0138540P.
PR 10-JUN-1999; 99US-0138847P.
PR 14-JUN-1999; 99US-0139119P.
PR 16-JUN-1999; 99US-0139452P.
PR 16-JUN-1999; 99US-0139453P.
PR 17-JUN-1999; 99US-0139492P.
PR 18-JUN-1999; 99US-0139454P.
PR 18-JUN-1999; 99US-0139455P.
PR 18-JUN-1999; 99US-0139456P.
PR 18-JUN-1999; 99US-0139457P.
PR 18-JUN-1999; 99US-0139458P.
PR 18-JUN-1999; 99US-0139459P.
PR 18-JUN-1999; 99US-0139460P.
PR 18-JUN-1999; 99US-0139461P.
PR 18-JUN-1999; 99US-0139462P.
PR 18-JUN-1999; 99US-0139463P.
PR 18-JUN-1999; 99US-0139750P.
PR 18-JUN-1999; 99US-0139763P.
PR 21-JUN-1999; 99US-0139817P.
PR 22-JUN-1999; 99US-0139899P.
PR 23-JUN-1999; 99US-0140353P.
PR 23-JUN-1999; 99US-0140354P.
PR 24-JUN-1999; 99US-0140695P.
PR 28-JUN-1999; 99US-0140823P.
PR 29-JUN-1999; 99US-0140991P.
PR 30-JUN-1999; 99US-0141287P.
PR 01-JUL-1999; 99US-0141842P.
PR 01-JUL-1999; 99US-0142154P.
PR 02-JUL-1999; 99US-0142055P.
PR 06-JUL-1999; 99US-0142390P.
PR 08-JUL-1999; 99US-0142803P.
PR 09-JUL-1999; 99US-0142920P.
PR 12-JUL-1999; 99US-0142977P.
PR 13-JUL-1999; 99US-0143542P.
PR 14-JUL-1999; 99US-0143624P.
PR 15-JUL-1999; 99US-0144005P.
PR 16-JUL-1999; 99US-0144085P.
PR 16-JUL-1999; 99US-0144086P.

PR 19-JUL-1999; 99US-0144325P.
PR 19-JUL-1999; 99US-0144331P.
PR 19-JUL-1999; 99US-0144332P.
PR 19-JUL-1999; 99US-0144333P.
PR 19-JUL-1999; 99US-0144334P.
PR 19-JUL-1999; 99US-0144335P.
PR 20-JUL-1999; 99US-0144352P.
PR 20-JUL-1999; 99US-0144632P.
PR 20-JUL-1999; 99US-0144884P.
PR 21-JUL-1999; 99US-0144814P.
PR 21-JUL-1999; 99US-0145086P.
PR 21-JUL-1999; 99US-0145088P.
PR 22-JUL-1999; 99US-0145085P.
PR 22-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145089P.
PR 22-JUL-1999; 99US-0145192P.
PR 23-JUL-1999; 99US-0145145P.
PR 23-JUL-1999; 99US-0145218P.
PR 23-JUL-1999; 99US-0145224P.
PR 26-JUL-1999; 99US-0145276P.
PR 27-JUL-1999; 99US-0145913P.
PR 27-JUL-1999; 99US-0145918P.
PR 27-JUL-1999; 99US-0145919P.
PR 28-JUL-1999; 99US-0145951P.
PR 02-AUG-1999; 99US-0146386P.
PR 02-AUG-1999; 99US-0146388P.
PR 02-AUG-1999; 99US-0146389P.
PR 03-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147204P.
PR 04-AUG-1999; 99US-0147302P.
PR 05-AUG-1999; 99US-0147192P.
PR 05-AUG-1999; 99US-0147260P.
PR 06-AUG-1999; 99US-0147303P.
PR 06-AUG-1999; 99US-0147416P.
PR 09-AUG-1999; 99US-0147493P.
PR 09-AUG-1999; 99US-0147935P.
PR 10-AUG-1999; 99US-0148171P.
PR 11-AUG-1999; 99US-0148319P.
PR 12-AUG-1999; 99US-0148341P.
PR 13-AUG-1999; 99US-0148565P.
PR 13-AUG-1999; 99US-0148684P.
PR 16-AUG-1999; 99US-0149368P.
PR 17-AUG-1999; 99US-0149175P.
PR 18-AUG-1999; 99US-0149426P.
PR 20-AUG-1999; 99US-0149722P.
PR 20-AUG-1999; 99US-0149723P.
PR 20-AUG-1999; 99US-0149929P.
PR 23-AUG-1999; 99US-0149902P.
PR 23-AUG-1999; 99US-0149930P.
PR 25-AUG-1999; 99US-0150566P.
PR 26-AUG-1999; 99US-0150884P.
PR 27-AUG-1999; 99US-0151065P.
PR 27-AUG-1999; 99US-0151066P.
PR 27-AUG-1999; 99US-0151080P.
PR 30-AUG-1999; 99US-0151303P.
PR 31-AUG-1999; 99US-0151438P.
PR 01-SEP-1999; 99US-0151930P.
PR 07-SEP-1999; 99US-0152363P.
PR 10-SEP-1999; 99US-0153070P.
PR 13-SEP-1999; 99US-0153758P.
PR 15-SEP-1999; 99US-0154018P.
PR 16-SEP-1999; 99US-0154039P.
PR 20-SEP-1999; 99US-0154779P.
PR 22-SEP-1999; 99US-0155139P.
PR 23-SEP-1999; 99US-0155486P.
PR 24-SEP-1999; 99US-0155659P.
PR 28-SEP-1999; 99US-0156458P.
PR 29-SEP-1999; 99US-0156596P.
PR 04-OCT-1999; 99US-0157117P.
PR 05-OCT-1999; 99US-0157753P.
PR 06-OCT-1999; 99US-0157865P.
PR 07-OCT-1999; 99US-0158029P.
PR 08-OCT-1999; 99US-0158232P.
PR 12-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0159293P.
PR 13-OCT-1999; 99US-0159294P.
PR 13-OCT-1999; 99US-0159295P.
PR 14-OCT-1999; 99US-0159329P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159637P.
PR 14-OCT-1999; 99US-0159638P.
PR 18-OCT-1999; 99US-0159584P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160768P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 21-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160989P.
PR 25-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161406P.
PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161992P.
PR 28-OCT-1999; 99US-0161993P.
PR 29-OCT-1999; 99US-0162142P.

Query Match 38.2%; Score 52; DB 3; Length 448;
Best Local Similarity 50.0%; Pred. No. 23;
Matches 12; Conservative 2; Mismatches 10; Indels 0; Gaps 0;

```
Qy          4 FSTEMANKAAEAVLKGQVETIVSF 27
              ||| :||     |||   ||| :|
Db        185 FSTRLANNLENLVLKEGPETIAAF 208
```

RESULT 23
AAG33444
ID AAG33444 standard; protein; 509 AA.
XX
AC AAG33444;
XX
DT 18-OCT-2000 (first entry)
XX
DE Zea mays protein fragment SEQ ID NO: 40523.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence; corn.
XX
OS Zea mays subsp. mays.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-00301439.
XX
PR 25-FEB-1999; 99US-0121825P.
PR 05-MAR-1999; 99US-0123180P.
PR 09-MAR-1999; 99US-0123548P.
PR 23-MAR-1999; 99US-0125788P.
PR 25-MAR-1999; 99US-0126264P.
PR 29-MAR-1999; 99US-0126785P.
PR 01-APR-1999; 99US-0127462P.
PR 06-APR-1999; 99US-0128234P.
PR 08-APR-1999; 99US-0128714P.
PR 16-APR-1999; 99US-0129845P.
PR 19-APR-1999; 99US-0130077P.
PR 21-APR-1999; 99US-0130449P.

PR 23-APR-1999; 99US-0130510P.
PR 23-APR-1999; 99US-0130891P.
PR 28-APR-1999; 99US-0131449P.
PR 30-APR-1999; 99US-0132048P.
PR 30-APR-1999; 99US-0132407P.
PR 04-MAY-1999; 99US-0132484P.
PR 05-MAY-1999; 99US-0132485P.
PR 06-MAY-1999; 99US-0132486P.
PR 06-MAY-1999; 99US-0132487P.
PR 07-MAY-1999; 99US-0132863P.
PR 11-MAY-1999; 99US-0134256P.
PR 14-MAY-1999; 99US-0134218P.
PR 14-MAY-1999; 99US-0134219P.
PR 14-MAY-1999; 99US-0134221P.
PR 14-MAY-1999; 99US-0134370P.
PR 18-MAY-1999; 99US-0134768P.
PR 19-MAY-1999; 99US-0134941P.
PR 20-MAY-1999; 99US-0135124P.
PR 21-MAY-1999; 99US-0135353P.
PR 24-MAY-1999; 99US-0135629P.
PR 25-MAY-1999; 99US-0136021P.
PR 27-MAY-1999; 99US-0136392P.
PR 28-MAY-1999; 99US-0136782P.
PR 01-JUN-1999; 99US-0137222P.
PR 03-JUN-1999; 99US-0137528P.
PR 04-JUN-1999; 99US-0137502P.
PR 07-JUN-1999; 99US-0137724P.
PR 08-JUN-1999; 99US-0138094P.
PR 10-JUN-1999; 99US-0138540P.
PR 10-JUN-1999; 99US-0138847P.
PR 14-JUN-1999; 99US-0139119P.
PR 16-JUN-1999; 99US-0139452P.
PR 16-JUN-1999; 99US-0139453P.
PR 17-JUN-1999; 99US-0139492P.
PR 18-JUN-1999; 99US-0139454P.
PR 18-JUN-1999; 99US-0139455P.
PR 18-JUN-1999; 99US-0139456P.
PR 18-JUN-1999; 99US-0139457P.
PR 18-JUN-1999; 99US-0139458P.
PR 18-JUN-1999; 99US-0139459P.
PR 18-JUN-1999; 99US-0139460P.
PR 18-JUN-1999; 99US-0139461P.
PR 18-JUN-1999; 99US-0139462P.
PR 18-JUN-1999; 99US-0139463P.
PR 18-JUN-1999; 99US-0139750P.
PR 18-JUN-1999; 99US-0139763P.
PR 21-JUN-1999; 99US-0139817P.
PR 22-JUN-1999; 99US-0139899P.
PR 23-JUN-1999; 99US-0140353P.
PR 23-JUN-1999; 99US-0140354P.
PR 24-JUN-1999; 99US-0140695P.
PR 28-JUN-1999; 99US-0140823P.
PR 29-JUN-1999; 99US-0140991P.
PR 30-JUN-1999; 99US-0141287P.
PR 01-JUL-1999; 99US-0141842P.
PR 01-JUL-1999; 99US-0142154P.
PR 02-JUL-1999; 99US-0142055P.
PR 06-JUL-1999; 99US-0142390P.
PR 08-JUL-1999; 99US-0142803P.
PR 09-JUL-1999; 99US-0142920P.
PR 12-JUL-1999; 99US-0142977P.
PR 13-JUL-1999; 99US-0143542P.
PR 14-JUL-1999; 99US-0143624P.
PR 15-JUL-1999; 99US-0144005P.
PR 16-JUL-1999; 99US-0144085P.
PR 16-JUL-1999; 99US-0144086P.
PR 19-JUL-1999; 99US-0144325P.
PR 19-JUL-1999; 99US-0144331P.
PR 19-JUL-1999; 99US-0144332P.
PR 19-JUL-1999; 99US-0144333P.
PR 19-JUL-1999; 99US-0144334P.
PR 19-JUL-1999; 99US-0144335P.
PR 20-JUL-1999; 99US-0144352P.
PR 20-JUL-1999; 99US-0144632P.
PR 20-JUL-1999; 99US-0144884P.
PR 21-JUL-1999; 99US-0144814P.
PR 21-JUL-1999; 99US-0145086P.
PR 21-JUL-1999; 99US-0145088P.
PR 22-JUL-1999; 99US-0145085P.
PR 22-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145089P.
PR 22-JUL-1999; 99US-0145192P.
PR 23-JUL-1999; 99US-0145145P.
PR 23-JUL-1999; 99US-0145218P.
PR 23-JUL-1999; 99US-0145224P.
PR 26-JUL-1999; 99US-0145276P.
PR 27-JUL-1999; 99US-0145913P.
PR 27-JUL-1999; 99US-0145918P.
PR 27-JUL-1999; 99US-0145919P.
PR 28-JUL-1999; 99US-0145951P.
PR 02-AUG-1999; 99US-0146386P.
PR 02-AUG-1999; 99US-0146388P.
PR 02-AUG-1999; 99US-0146389P.
PR 03-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147204P.
PR 04-AUG-1999; 99US-0147302P.
PR 05-AUG-1999; 99US-0147192P.
PR 05-AUG-1999; 99US-0147260P.
PR 06-AUG-1999; 99US-0147303P.
PR 06-AUG-1999; 99US-0147416P.
PR 09-AUG-1999; 99US-0147493P.
PR 09-AUG-1999; 99US-0147935P.
PR 10-AUG-1999; 99US-0148171P.
PR 11-AUG-1999; 99US-0148319P.
PR 12-AUG-1999; 99US-0148341P.
PR 13-AUG-1999; 99US-0148565P.
PR 13-AUG-1999; 99US-0148684P.
PR 16-AUG-1999; 99US-0149368P.
PR 17-AUG-1999; 99US-0149175P.
PR 18-AUG-1999; 99US-0149426P.
PR 20-AUG-1999; 99US-0149722P.
PR 20-AUG-1999; 99US-0149723P.
PR 20-AUG-1999; 99US-0149929P.
PR 23-AUG-1999; 99US-0149902P.
PR 23-AUG-1999; 99US-0149930P.
PR 25-AUG-1999; 99US-0150566P.
PR 26-AUG-1999; 99US-0150884P.
PR 27-AUG-1999; 99US-0151065P.
PR 27-AUG-1999; 99US-0151066P.
PR 27-AUG-1999; 99US-0151080P.
PR 30-AUG-1999; 99US-0151303P.
PR 31-AUG-1999; 99US-0151438P.
PR 01-SEP-1999; 99US-0151930P.
PR 07-SEP-1999; 99US-0152363P.
PR 10-SEP-1999; 99US-0153070P.
PR 13-SEP-1999; 99US-0153758P.
PR 15-SEP-1999; 99US-0154018P.
PR 16-SEP-1999; 99US-0154039P.
PR 20-SEP-1999; 99US-0154779P.
PR 22-SEP-1999; 99US-0155139P.
PR 23-SEP-1999; 99US-0155486P.
PR 24-SEP-1999; 99US-0155659P.
PR 28-SEP-1999; 99US-0156458P.
PR 29-SEP-1999; 99US-0156596P.
PR 04-OCT-1999; 99US-0157117P.
PR 05-OCT-1999; 99US-0157753P.
PR 06-OCT-1999; 99US-0157865P.
PR 07-OCT-1999; 99US-0158029P.
PR 08-OCT-1999; 99US-0158232P.
PR 12-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0159293P.
PR 13-OCT-1999; 99US-0159294P.
PR 13-OCT-1999; 99US-0159295P.
PR 14-OCT-1999; 99US-0159329P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.

```
PR   14-OCT-1999;   99US-0159637P.
PR   14-OCT-1999;   99US-0159638P.
PR   18-OCT-1999;   99US-0159584P.
PR   21-OCT-1999;   99US-0160741P.
PR   21-OCT-1999;   99US-0160767P.
PR   21-OCT-1999;   99US-0160768P.
PR   21-OCT-1999;   99US-0160770P.
PR   21-OCT-1999;   99US-0160814P.
PR   21-OCT-1999;   99US-0160815P.
PR   22-OCT-1999;   99US-0160980P.
PR   22-OCT-1999;   99US-0160981P.
PR   22-OCT-1999;   99US-0160989P.
PR   25-OCT-1999;   99US-0161404P.
PR   25-OCT-1999;   99US-0161405P.
PR   25-OCT-1999;   99US-0161406P.
PR   26-OCT-1999;   99US-0161359P.
PR   26-OCT-1999;   99US-0161360P.
PR   26-OCT-1999;   99US-0161361P.
PR   28-OCT-1999;   99US-0161920P.
PR   28-OCT-1999;   99US-0161992P.
PR   28-OCT-1999;   99US-0161993P.
PR   29-OCT-1999;   99US-0162142P.

  Query Match             38.2%;  Score 52;  DB 3;  Length 509;
  Best Local Similarity   50.0%;  Pred. No. 27;
  Matches   12;  Conservative    2;  Mismatches   10;  Indels    0;  Gaps    0;

Qy          4 FSTEMANKAAEAVLKGQVETIVSF 27
              ||| :||     |||   ||| :|
Db        246 FSTRLANNLENLVLKEGPETIAAF 269


RESULT 24
ABB60387
ID   ABB60387 standard; protein; 1049 AA.
XX
AC   ABB60387;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Drosophila melanogaster polypeptide SEQ ID NO 7953.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US009231.
XX
PR   23-MAR-2000; 2000US-0191637P.
PR   11-JUL-2000; 2000US-00614150.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Venter JC,  Adams M,  Li PWD,  Myers EW;
XX
DR   WPI; 2001-656860/75.
DR   N-PSDB; ABL04490.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signaling and cell-cell
PT   interactions.
XX
PS   Disclosure; SEQ ID NO 7953; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-
CC   ABB72072). The sequence data for this patent did not form part of the
CC   printed specification, but was obtained in electronic format directly
CC   from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 1049 AA;

  Query Match             37.5%;  Score 51;  DB 4;  Length 1049;
  Best Local Similarity   40.7%;  Pred. No. 92;
  Matches   11;  Conservative    5;  Mismatches   11;  Indels    0;  Gaps    0;

Qy          2 YVFSTEMANKAAEAVLKGQVETIVSFH 28
              ||  ::||  | |||  |::: |   |
Db        471 YVSCSDMAASATEAVRSGELKIIPEHH 497


RESULT 25
ABB89793
ID   ABB89793 standard; protein; 130 AA.
XX
AC   ABB89793;
XX
DT   24-MAY-2002  (first entry)
XX
DE   Human polypeptide SEQ ID NO 2169.
XX
KW   Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;
KW   antiallergic; hepatotropic; antidiabetic; antiinflammatory; antiulcer;
KW   vulnerary; anticonvulsant; antibacterial; antifungal; antiparasitic;
KW   cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;
KW   neurological disease; infection; human; secreted protein.
XX
OS   Homo sapiens.
XX
PN   WO200190304-A2.
XX
PD   29-NOV-2001.
XX
PF   18-MAY-2001; 2001WO-US016450.
XX
PR   19-MAY-2000; 2000US-0205515P.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Birse CE,  Rosen CA;
XX
DR   WPI; 2002-122018/16.
DR   N-PSDB; ABL90202.
XX
PT   Novel 1405 isolated polypeptides, useful for diagnosis, treatment and
PT   prevention of neural, immune system, muscular, reproductive,
PT   gastrointestinal, pulmonary, cardiovascular, renal and proliferative
PT   disorders.
XX
PS   Claim 11; SEQ ID NO 2169; 2081pp + Sequence Listing; English.
XX
CC   The invention relates to novel genes (ABL89449-ABL90853) and proteins
CC   (ABB89040-ABB90444) useful for preventing, treating or ameliorating
CC   medical conditions e.g. by protein or gene therapy. The genes are
CC   isolated from a range of human tissues disclosed in the specification.
CC   The nucleic acids, proteins, antibodies and (ant)agonists are useful in
CC   the diagnosis, treatment and prevention of: (a) cancer, e.g. breast and
CC   ovarian cancer and other cancers of the adrenal gland, bone, bone marrow,
CC   breast, gastrointestinal tract, liver, lung, or urogenital; (b) immune
CC   disorders e.g. Addison's disease, allergies, autoimmune haemolytic
CC   anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's disease,
CC   multiple sclerosis, rheumatoid arthritis and ulcerative colitis; (c)
CC   cardiovascular disorders such as myocardial ischaemias; (d) wound healing
CC   ; (e) neurological diseases e.g. cerebral anoxia and epilepsy; and (f)
CC   infectious diseases such as viral, bacterial, fungal and parasitic
CC   infections. Note: The sequence data for this patent did not form part of
```

```
CC   the printed specification, but was obtained in electronic format directly
CC   from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 130 AA;

  Query Match             36.8%;  Score 50;  DB 5;  Length 130;
  Best Local Similarity   44.4%;  Pred. No. 11;
  Matches   12;  Conservative    4;  Mismatches   11;  Indels    0;  Gaps    0;

Qy          2 YVFSTEMANKAAEAVLKGQVETIVSFH 28
              :|   ||  :||:||  | :|   |||
Db         67 FVRCQEMGARAAKAVESGALELSPSFH 93


RESULT 26
ABG15088
ID   ABG15088 standard; protein; 365 AA.
XX
AC   ABG15088;
XX
DT   18-FEB-2002  (first entry)
XX
DE   Novel human diagnostic protein #15079.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS   Homo sapiens.
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US008631.
XX
PR   31-MAR-2000; 2000US-00540217.
PR   23-AUG-2000; 2000US-00649167.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Drmanac RT,  Liu C,  Tang YT;
XX
DR   WPI; 2001-639362/73.
DR   N-PSDB; AAS79275.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity.
XX
PS   Claim 20; SEQ ID NO 45447; 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and polypeptide (II)
CC   sequences. (I) is useful as hybridisation probes, polymerase chain
CC   reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
CC   and in recombinant production of (II). The polynucleotides are also used
CC   in diagnostics as expressed sequence tags for identifying expressed
CC   genes. (I) is useful in gene therapy techniques to restore normal
CC   activity of (II) or to treat disease states involving (II). (II) is
CC   useful for generating antibodies against it, detecting or quantitating a
CC   polypeptide in tissue, as molecular weight markers and as a food
CC   supplement. (II) and its binding partners are useful in medical imaging
CC   of sites expressing (II). (I) and (II) are useful for treating disorders
CC   involving aberrant protein expression or biological activity. The
CC   polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic
CC   amino acid sequences of the invention. Note: The sequence data for this
CC   patent did not appear in the printed specification, but was obtained in
CC   electronic format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 365 AA;

  Query Match             36.8%;  Score 50;  DB 4;  Length 365;
  Best Local Similarity   29.6%;  Pred. No. 38;
  Matches    8;  Conservative    8;  Mismatches   11;  Indels    0;  Gaps    0;

Qy          2 YVFSTEMANKAAEAVLKGQVETIVSFH 28
              |||:  : | :|: :: |  | ::  |
Db         67 YVFAYPVCNASAKTIISGLTECLIHCH 93


RESULT 27
AAE15740
ID   AAE15740 standard; protein; 621 AA.
XX
AC   AAE15740;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Human aminoacyl-tRNA synthetase-3 (ATRS-3).
XX
KW   Human; aminoacyl-tRNA synthetase; ATRS; cell proliferative disorder;
KW   cancer; actinic keratosis; arteriosclerosis; atherosclerosis; bursitis;
KW   cirrhosis; immunostimulant; antithyroid; immunosuppressive; fungicide;
KW   hepatitis; psoriasis; autoimmune disorder; inflammatory disorder;
KW   acquired immune deficiency syndrome; AIDS; Addison's disease; allergy;
KW   adult respiratory distress syndrome; anaemia; autoimmune thyroiditis;
KW   osteoporosis; autoimmune haemolytic anaemia; hepatotropic; antihelmintic;
KW   Crohn's disease; atopic dermatitis; diabetic mellitus; Graves' disease;
KW   glomerulonephritis; rheumatoid arthritis; scleroderma; osteopathic;
KW   systemic lupus erythematosus; systemic sclerosis; ulcerative colitis;
KW   haemodialysis; uveitis; infection; single nucleotide polymorphism;
KW   gene therapy; cytostatic; dermatological; antiulcer; antibacterial;
KW   virucide; antiparasitic; protozoacide; tranquilliser; vulnerary;
KW   human immunodeficiency virus; antiinflammatory; nephrotropic;
KW   ophthalmological; anti-HIV; asthma.
XX
OS   Homo sapiens.
XX
PN   WO200190330-A2.
XX
PD   29-NOV-2001.
XX
PF   22-MAY-2001; 2001WO-US016808.
XX
PR   25-MAY-2000; 2000US-0207248P.
PR   01-JUN-2000; 2000US-0208791P.
PR   08-JUN-2000; 2000US-0210585P.
XX
PA   (INCY-) INCYTE GENOMICS INC.
XX
PI   Yue H,  Tang TY,  Patterson C,  Gandhi AR,  Tribouley CM,  Lee EA;
PI   Yao MG,  Bandman O,  Lu DAM;
XX
DR   WPI; 2002-083106/11.
DR   N-PSDB; AAD25342.
XX
PT   Novel human aminoacyl-tRNA synthetase polypeptides and polynucleotides
PT   for diagnosing, preventing or treating Addison's disease, allergies,
PT   asthma, rheumatoid arthritis, scleroderma, systemic lupus erythematosus.
XX
PS   Claim 1; Page 98-99; 103pp; English.
XX
CC   The invention relates to human aminoacyl-tRNA synthetases (ATRS) and
CC   their corresponding nucleic acids. ATRS is useful in screening for a
CC   compound that modulates the activity of the polypeptide or that binds to
CC   the polypeptide. ATRS is also useful as an immunogen for preparing
CC   polyclonal or monoclonal antibodies by hybridoma technology. The agonist
CC   and antagonist of ATRS are useful for treating a disease or condition
CC   associated with decreased or overexpression of functional ATRS in a
CC   patient. ATRS, its DNA and its modulators are useful for diagnosis,
CC   treatment and prevention of cell proliferative disorders such as cancer,
```

```
CC   actinic keratosis, arteriosclerosis, atherosclerosis, bursitis,
CC   cirrhosis, hepatitis and psoriasis, autoimmune/inflammatory disorders
CC   such as acquired immune deficiency syndrome (AIDS), adult respiratory
CC   distress syndrome, Addison's disease, allergies, anaemia, asthma,
CC   osteoporosis, autoimmune haemolytic anaemia, autoimmune thyroiditis,
CC   Crohn's disease, atopic dermatitis, diabetic mellitus, Graves' disease,
CC   glomerulonephritis, rheumatoid arthritis, scleroderma, systemic lupus
CC   erythematosus, systemic sclerosis, ulcerative colitis, haemodialysis,
CC   uveitis, viral, bacterial, fungal, parasitic, protozoal, helminthic
CC   infections and trauma. ATRS DNA is also useful for generating
CC   hybridisation probes useful in mapping the naturally occurring genomic
CC   sequence and oligonucleotide primers derived from it are useful to detect
CC   single nucleotide polymorphisms. ATRS DNA is used in gene therapy. The
CC   present sequence is human ATRS-3 protein
XX
SQ   Sequence 621 AA;

  Query Match             36.8%;  Score 50;  DB 5;  Length 621;
  Best Local Similarity   44.4%;  Pred. No. 71;
  Matches   12;  Conservative    4;  Mismatches   11;  Indels    0;  Gaps    0;

Qy          2 YVFSTEMANKAAEAVLKGQVETIVSFH 28
              :|   ||  :||:||  | :|   |||
Db         39 FVRCQEMGARAAKAVESGALELSPSFH 65


RESULT 28
AAB43285
ID   AAB43285 standard; protein; 631 AA.
XX
AC   AAB43285;
XX
DT   08-FEB-2001  (first entry)
XX
DE   Human ORFX ORF3049 polypeptide sequence SEQ ID NO:6098.
XX
KW   Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;
KW   vulnerary; antipsoriatic; antiparkinsonian; nootropic; neuroprotective;
KW   anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;
KW   immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
KW   hypotensive; dermatological; immunosuppressive; antiinflammatory;
KW   antiviral; antibacterial; antifungal; antirheumatic; antithyroid;
KW   antianaemic; gene therapy; cancer; proliferative disorder; hypertension;
KW   neurodegenerative disorder; osteoarthritis; graft vs host disease;
KW   cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
KW   cholesterol ester storage; systemic lupus erythematosus; infection;
KW   severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
KW   allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
KW   bone damage; cartilage damage; antiinflammatory disease; coagulation;
KW   thrombosis; contraceptive.
XX
OS   Homo sapiens.
XX
PN   WO200058473-A2.
XX
PD   05-OCT-2000.
XX
PF   31-MAR-2000; 2000WO-US008621.
XX
PR   31-MAR-1999;   99US-0127607P.
PR   02-APR-1999;   99US-0127636P.
PR   05-APR-1999;   99US-0127728P.
PR   30-MAR-2000; 2000US-00540763.
XX
PA   (CURA-) CURAGEN CORP.
XX
PI   Shimkets RA,  Leach M;
XX
DR   WPI; 2000-602362/57.
DR   N-PSDB; AAC77494.
XX
PT   Novel nucleic acids and peptides derived from open reading frame X,
PT   useful for treating e.g. cancers, proliferative disorders,
PT   neurodegenerative disorders and cardiovascular disease.
XX
PS   Claim 11; Page 5281-5283; 5507pp; English.
XX
CC   AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
CC   which represent the human ORFX open reading frames 1 to 3161. The ORFX
CC   sequences have activities such as: cytostatic; hepatotropic; vulnerary;
CC   antipsoriatic; antiparkinsonian; nootropic; neuroprotective; osteopathic;
CC   anticonvulsant; antiarthritic; immunosuppressant; immunostimulant;
CC   cardiant; thrombolytic; coagulant; vasotropic; antidiabetic; hypotensive;
CC   dermatological; immunosuppressive; antiinflammatory; antibacterial;
CC   antiviral; antifungal; antirheumatic; antithyroid; and antianaemic. The
CC   sequences can be used for determining the presence of or predisposition
CC   to, or preventing or treating pathological conditions associated with an
CC   ORFX-associated disorder. The nucleic acids can be used to express ORFX
CC   proteins in gene therapy vectors. The proteins and nucleic acids may be
CC   used to treat cancers, proliferative disorders, neurodegenerative
CC   disorders, osteoarthritis, graft vs host disease, cardiovascular disease,
CC   diabetes mellitus, hypertension, hypothyroidism, cholesterol ester
CC   storage, systemic lupus erythematosus, severe combined immunodeficiency
CC   (SCID), AIDS, viral, bacterial or fungal infection, malaria, autoimmune
CC   disorders, asthma, allergies, aplastic anaemia, burns, wounds, bone and
CC   cartilage damage, nocturnal haemoglobinuria, antiinflammatory disease; to
CC   enhance coagulation; to inhibit thrombosis; and as a contraceptive
XX
SQ   Sequence 631 AA;

  Query Match             36.8%;  Score 50;  DB 3;  Length 631;
  Best Local Similarity   44.4%;  Pred. No. 72;
  Matches   12;  Conservative    4;  Mismatches   11;  Indels    0;  Gaps    0;

Qy          2 YVFSTEMANKAAEAVLKGQVETIVSFH 28
              :|   ||  :||:||  | :|   |||
Db         47 FVRCQEMGARAAKAVESGALELSPSFH 73


RESULT 29
ABM80775
ID   ABM80775 standard; protein; 694 AA.
XX
AC   ABM80775;
XX
DT   18-NOV-2004  (first entry)
XX
DE   Tumour-associated antigenic target (TAT) polypeptide PRO81404, SEQ:1996.
XX
KW   Tumour-associated antigenic target; TAT; human; overexpression; cancer;
KW   tumour; diagnosis; cell proliferative disorder; breast cancer;
KW   colorectal cancer; lung cancer; ovarian cancer; liver cancer;
KW   central nervous system cancer; bladder cancer; pancreatic cancer;
KW   cervical cancer; melanoma; leukaemia; hybridisation probe;
KW   chromosome identification; chromosome mapping; gene mapping;
KW   gene therapy; cytostatic.
XX
OS   Homo sapiens.
XX
PN   WO2004030615-A2.
XX
PD   15-APR-2004.
XX
PF   29-SEP-2003; 2003WO-US028547.
XX
PR   02-OCT-2002; 2002US-0414971P.
XX
PA   (GETH ) GENENTECH INC.
XX
PI   Wu TD,  Zhang Z,  Zhou Y;
XX
DR   WPI; 2004-347921/32.
DR   N-PSDB; ACN38464.
XX
PT   New tumor-associated antigenic target polypeptides and nucleic acids,
PT   useful in preparing a medicament for treating or detecting a
```

```
PT   proliferative disorder, e.g. breast, lung, colorectal, ovarian or
PT   prostate cancer or tumor.
XX
PS   Claim 12; SEQ ID NO 1996; 7273pp; English.
XX
CC   The invention relates to human tumour-associated antigenic target (TAT)
CC   polypeptides, and their related nucleic acids. The TAT polypeptides are
CC   overexpressed in cancer tissues compared to normal tissues, and may thus
CC   serve as effective targets for the diagnosis and treatment of cancer in
CC   mammals. The invention also relates to nucleic acid and polypeptide
CC   sequences at least 80% identical to the TAT nucleic acids and
CC   polypeptides; expression vectors and host cells comprising a TAT nucleic
CC   acid; an antibody specific for a TAT polypeptide; a peptide or organic
CC   molecule which binds to a TAT polypeptide; fusion proteins comprising a
CC   TAT polypeptide; and methods and compositions for the treatment or
CC   diagnosis of cancer in mammals. TAT polypeptides, nucleic acids,
CC   antibodies, antagonists, binding molecules and compositions are useful
CC   for diagnosing or treating a cell proliferative disorder associated with
CC   increased TAT expression, particularly cancers such as breast cancer,
CC   colorectal cancer, lung cancer, ovarian cancer, liver cancer, bladder
CC   cancer, pancreatic cancer, cervical cancer, cancers of the central
CC   nervous system, melanoma and leukaemia. TAT nucleic acids may further be
CC   used as hybridisation probes, in chromosome and gene mapping, in
CC   chromosome identification and in gene therapy. The present sequence
CC   represents a TAT polypeptide of the invention
XX
SQ   Sequence 694 AA;

  Query Match             36.8%;  Score 50;  DB 8;  Length 694;
  Best Local Similarity   44.4%;  Pred. No. 81;
  Matches   12;  Conservative    4;  Mismatches   11;  Indels    0;  Gaps    0;

Qy          2 YVFSTEMANKAAEAVLKGQVETIVSFH 28
              :|   ||  :||:||  | :|   |||
Db        112 FVRCQEMGARAAKAVESGALELSPSFH 138


RESULT 30
ADI16244
ID   ADI16244 standard; protein; 961 AA.
XX
AC   ADI16244;
XX
DT   22-APR-2004  (first entry)
XX
DE   Human nucleic acid-associated protein (NAAP) #29.
XX
KW   human; nucleic acid-associated protein; NAAP; autoimmune disorder;
KW   inflammatory disorder; AIDS; allergy; infection; metabolic disorder;
KW   obesity; reproductive disorder; infertility; neurological disorder;
KW   Parkinson's disease; Alzheimer's disease; cardiovascular disorder;
KW   myocardial infarction; hypertension; eye disorder;
KW   cell proliferative disease; cancer.
XX
OS   Homo sapiens.
XX
PN   WO2003094848-A2.
XX
PD   20-NOV-2003.
XX
PF   09-MAY-2003; 2003WO-US014450.
XX
PR   10-MAY-2002; 2002US-0379843P.
PR   24-MAY-2002; 2002US-0383457P.
PR   31-MAY-2002; 2002US-0384699P.
PR   06-JUN-2002; 2002US-0387265P.
XX
PA   (INCY-) INCYTE CORP.
XX
PI   Kable AE,  Elliott VS,  Tran UK,  Ramkumar J,  Marquis JP,  Chawla NK;
PI   Richardson TW,  Bulloch SA,  Khare R,  Lee SY,  Lal PG,  Tang YT,  Yue H;
PI   Swarnakar A,  Becha SD,  Hafalia AJA,  Chang H,  Baughn MR,  Borowsky ML;
PI   Gietzen KJ,  He A,  Forsythe IJ,  Sprague WW,  Blake JJ,  Warren BA;
PI   Mason PM,  Ison CH,  Lindquist EA,  Wilson AD,  Jin P;
XX
DR   WPI; 2004-011999/01.
DR   N-PSDB; ADI16294.
XX
PT   New human nucleic acid associated proteins and polynucleotides, useful
PT   for diagnosing, preventing or treating diseases or conditions associated
PT   with aberrant protein expression, e.g. cancer, AIDS, atherosclerosis or
PT   stroke.
XX
PS   Claim 1; SEQ ID NO 29; 400pp; English.
XX
CC   The invention comprises the amino acid and coding sequences of human
CC   nucleic acid-associated proteins (NAAP). The DNA and protein sequences of
CC   the invention are useful in diagnosing, preventing and treating
CC   diseases/conditions associated with altered expression of NAAP, such as:
CC   autoimmune/inflammatory disorders (e.g. AIDS and allergies), infections
CC   (e.g. bacterial and viral), metabolic disorders (e.g. obesity),
CC   reproductive disorders (e.g. infertility), neurological disorders (e.g.
CC   Parkinson's disease and Alzheimer's disease), cardiovascular disorders
CC   (e.g. myocardial infarction and hypertension), eye disorders, or cell
CC   proliferative diseases (e.g. cancer). The present amino acid sequence
CC   represents a human NAAP protein of the invention.
XX
SQ   Sequence 961 AA;

  Query Match             36.8%;  Score 50;  DB 8;  Length 961;
  Best Local Similarity   44.4%;  Pred. No. 1.2e+02;
  Matches   12;  Conservative    4;  Mismatches   11;  Indels    0;  Gaps    0;

Qy          2 YVFSTEMANKAAEAVLKGQVETIVSFH 28
              :|   ||  :||:||  | :|   |||
Db        481 FVRCQEMGARAAKAVESGALELSPSFH 507


RESULT 31
ABB08919
ID   ABB08919 standard; protein; 1063 AA.
XX
AC   ABB08919;
XX
DT   21-OCT-2002  (first entry)
XX
DE   Human aminoacyl tRNA synthetase (ATRS-1).
XX
KW   Human; aminoacyl tRNA synthetase; ATRS-1; valyl-tRNA synthetase; Class I;
KW   Rossman fold; cell proliferative disorder; cancer; psoriasis;
KW   atherosclerosis; cirrhosis; hepatitis; autoimmune disorder;
KW   inflammatory disorder; allergy; acquired immunodeficiency syndrome; AIDS;
KW   anaemia; diabetes; dermatomyositis; polymyositis; rheumatoid arthritis;
KW   trauma; infection; immunomodulator; immunosuppressive; cytostatic;
KW   antiarthritic; gene therapy; enzyme.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Region          1. .612
FT                   /note= "Cytosolic N-terminal region"
FT   Domain          112. .794
FT                   /note= "Aminoacyl tRNA synthetase Class I (I, L, M and V)
FT                   domain"
FT   Region          132. .181
FT                   /note= "Aminoacyl tRNA synthetase Class I signature
FT                   motif"
FT   Region          139. .150
FT                   /note= "Valyl-tRNA synthetase signature motif"
FT   Region          146. .157
FT                   /note= "Aminoacyl tRNA synthetase Class I signature
FT                   motif"
FT   Region          351. .368
FT                   /note= "Valyl-tRNA synthetase signature motif"
FT   Region          467. .480
FT                   /note= "Valyl-tRNA synthetase signature motif"
```

```
FT   Region          579. .600
FT                   /note= "Valyl-tRNA synthetase signature motif"
FT   Region          610. .628
FT                   /note= "Valyl-tRNA synthetase signature motif"
FT   Domain          613. .634
FT                   /label= Transmembrane_domain
FT   Domain          841. .858
FT                   /label= Transmembrane_domain
XX
PN   WO200259323-A2.
XX
PD   01-AUG-2002.
XX
PF   13-DEC-2001; 2001WO-US048575.
XX
PR   15-DEC-2000; 2000US-0255963P.
XX
PA   (INCY-) INCYTE GENOMICS INC.
XX
PI   Lee EA,  Baughn MR;
XX
DR   WPI; 2002-599795/64.
DR   N-PSDB; ABA97729.
XX
PT   New aminoacyl tRNA synthetases, useful for diagnosing, treating or
PT   preventing autoimmune or inflammatory disorders (e.g. AIDS, allergies or
PT   anemia) or cell proliferative disorders (e.g. cancers, atherosclerosis or
PT   hepatitis).
XX
PS   Claim 1; Page 89-91; 92pp; English.
XX
CC   The invention relates to a novel human aminoacyl tRNA synthetase
CC   designated ATRS-1 (ABB08919) and cDNA encoding it (ABA97729). ATRS-1 is
CC   thought to be a valyl-tRNA synthetase, based on its 50% homology to mouse
CC   valyl-tRNA synthetase and the presence of a tRNA synthetase Class I (I,
CC   L, M and V) domain. Class I enzymes such as ATRS-1 contain a catalytic
CC   domain based on a nucleotide-binding motif known as the Rossman fold, and
CC   add amino acids to the 2' hydroxyl group at the 3' end of tRNAs. ATRS-1
CC   nucleotides, polypeptides, agonists and antagonists may be used for
CC   diagnosing, treating or preventing disorders associated with aberrant
CC   expression of aminoacyl tRNA synthetases. Such disorders include cell
CC   proliferative disorders (e.g., cancers, psoriasis, atherosclerosis,
CC   cirrhosis and hepatitis); autoimmune or inflammatory disorders (e.g.,
CC   allergies, AIDS (acquired immunodeficiency syndrome), anaemia, diabetes,
CC   dermatomyositis, polymyositis and rheumatoid arthritis); trauma; and
CC   viral, bacterial, fungal, parasitic, protozoal or helminthic infections.
CC   They are also useful in screening for modulators of ATRS-1 expression or
CC   activity. The present sequence represents human aminoacyl tRNA synthetase
CC   ATRS-1
XX
SQ   Sequence 1063 AA;

  Query Match             36.8%;  Score 50;  DB 5;  Length 1063;
  Best Local Similarity   44.4%;  Pred. No. 1.3e+02;
  Matches   12;  Conservative    4;  Mismatches   11;  Indels    0;  Gaps    0;

Qy          2 YVFSTEMANKAAEAVLKGQVETIVSFH 28
              :|   ||  :||:||  | :|   |||
Db        481 FVRCQEMGARAAKAVESGALELSPSFH 507


RESULT 32
ADJ70652
ID   ADJ70652 standard; protein; 1063 AA.
XX
AC   ADJ70652;
XX
DT   06-MAY-2004  (first entry)
XX
DE   Human heat mitochondrial protein as a therapeutic target SeqID2458.
XX
KW   mitochondrial; human; screening assay; diabetes mellitus;
KW   Huntington's disease; osteoarthritis;
KW   Leber's hereditary optic neuropathy; LHON;
KW   mitochondrial encephalopathy lactic acidosis and stroke; MELAS;
KW   myoclonic epilepsy ragged red fibre syndrome; MERRF; cancer;
KW   neuroprotective; nootropic; antidiabetic; anticonvulsant; antiarthritic;
KW   osteopathic; ophthalmological; cytostatic.
XX
OS   Homo sapiens.
XX
PN   WO2003087768-A2.
XX
PD   23-OCT-2003.
XX
PF   04-APR-2003; 2003WO-US010870.
XX
PR   12-APR-2002; 2002US-0372843P.
PR   17-JUN-2002; 2002US-0389987P.
PR   20-SEP-2002; 2002US-0412418P.
XX
PA   (MITO-) MITOKOR.
PA   (BUCK-) BUCK INST AGE RES.
XX
PI   Ghosh SS,  Fahy ED,  Zhang B,  Gibson BW,  Taylor SW,  Glenn GM;
PI   Warnock DE;
XX
DR   WPI; 2003-845369/78.
XX
PT   Identifying a mitochondrial target for drug screening assays and for
PT   treating diseases associated with altered mitochondrial function,
PT   comprises detecting a modified polypeptide in a sample and correlating
PT   with the disease.
XX
PS   Claim 1; SEQ ID NO 2458; 180pp; English.
XX
CC   This invention relates to novel mitochondrial targets that can be used
CC   for therapeutic intervention in treating a disease associated with
CC   altered mitochondrial function. Specifically, it refers to a method for
CC   identifying proteins of the human heart mitochondrial proteome that are
CC   useful for drug screening assays, as well as therapeutic targets. The
CC   present invention describes a method for identifying such proteins that
CC   can be used in the treatment of various diseases associated with altered
CC   mitochondrial function including diabetes mellitus, Huntington's disease,
CC   osteoarthritis, Leber's hereditary optic neuropathy (LHON), mitochondrial
CC   encephalopathy lactic acidosis and stroke (MELAS), myoclonic epilepsy
CC   ragged red fibre syndrome (MERRF) or cancer. Accordingly, these
CC   compositions have neuroprotective, nootropic, antidiabetic,
CC   anticonvulsant, antiarthritic, osteopathic, ophthalmological and
CC   cytostatic activities. This polypeptide sequence is a human heart
CC   mitochondrial protein of the invention.
XX
SQ   Sequence 1063 AA;

  Query Match             36.8%;  Score 50;  DB 7;  Length 1063;
  Best Local Similarity   44.4%;  Pred. No. 1.3e+02;
  Matches   12;  Conservative    4;  Mismatches   11;  Indels    0;  Gaps    0;

Qy          2 YVFSTEMANKAAEAVLKGQVETIVSFH 28
              :|   ||  :||:||  | :|   |||
Db        481 FVRCQEMGARAAKAVESGALELSPSFH 507


RESULT 33
ADK40961
ID   ADK40961 standard; protein; 1078 AA.
XX
AC   ADK40961;
XX
DT   06-MAY-2004  (first entry)
XX
DE   Novel human kinase protein #68.
XX
KW   cytostatic; immunomodulator; cardiant; neuroprotective; nootropic;
KW   antiparkinsonian; virucide; antibacterial; fungicide; ophthalmological;
KW   analgesic; hypotensive; immunosuppressive; kinase inhibitor; kinase;
```

KW cancer; peripheral nervous system; central nervous system;
KW Alzheimer's disease; Parkinson's disease; multiple sclerosis;
KW amyotrophic lateral sclerosis; viral infection; prion infection;
KW ocular disease; migraine; pain; sexual dysfunction; mood disorder;
KW attention disorder; cognition disorder; hypotension; hypertension;
KW psychotic disorder; neurological disorder; dyskinesia;
KW metabolic disorder; organ transplant rejection; enzyme.
XX
OS Homo sapiens.
XX
PN WO2003057841-A2.
XX
PD 17-JUL-2003.
XX
PF 31-DEC-2002; 2002WO-US041687.
XX
PR 31-DEC-2001; 2001US-0343169P.
XX
PA (GRIG/) GRIGORIEV I V.
PA (SUDA/) SUDARSANAM S.
XX
PI Grigoriev IV, Sudarsanam S;
XX
DR WPI; 2003-587115/55.
XX
PT New isolated, enriched or purified nucleic acid molecule encoding a
PT kinase polypeptide, useful for treating cancer, immune-related diseases,
PT cardiovascular disease, brain or neuronal-associated diseases and
PT metabolic disorders.
XX
PS Claim 1; SEQ ID NO 68; 491pp; English.
XX
CC The invention elates to novel isolated, enriched or purified nucleic acid
CC molecules encoding a kinase polypeptide. The nucleic acid molecule
CC comprises a sequence that: (a) encodes a kinase polypeptide; (b) is a
CC complement of (a); (c) hybridizes under stringent conditions to (a) and
CC encodes a naturally occurring kinase polypeptide; (d) encodes the
CC polypeptide in (a), except that it lacks one or more, but not all, of an
CC N-terminal domain, C-terminal catalytic domain, a catalytic domain, a C-
CC terminal domain, a coiled-coil structure region, a spacer region and a C-
CC terminal tail; or (e) is a complement of (d). The nucleic acid molecules,
CC polypeptides, methods and substance are useful for treating cancers,
CC immune-related diseases or disorders, cardiovascular disease, brain or
CC neuronal-associated diseases, and metabolic disorders. The disorders are
CC preferably cancers of the tissues or of hematopoietic origin, diseases of
CC the central or peripheral nervous system, Alzheimer's disease,
CC Parkinson's disease, multiple sclerosis, amyotrophic lateral sclerosis,
CC viral infections, infections caused by prions, infections caused by
CC bacteria, infections caused by fungi, ocular diseases, migraines, pain,
CC sexual dysfunction, mood disorders, attention disorders, cognition
CC disorders, hypotension, hypertension, psychotic disorders, neurological
CC disorders, dyskinesias, metabolic disorders and organ transplant
CC rejection. This sequence corresponds to one of the kinase polypeptides of
CC the invention.
XX
SQ Sequence 1078 AA;

```
  Query Match             36.8%;  Score 50;  DB 7;  Length 1078;
  Best Local Similarity   44.4%;  Pred. No. 1.4e+02;
  Matches   12;  Conservative    4;  Mismatches   11;  Indels    0;  Gaps    0;

Qy          2 YVFSTEMANKAAEAVLKGQVETIVSFH 28
              :|   ||  :||:||  | :|   |||
Db        469 FVRCQEMGARAAKAVESGALELSPSFH 495
```

RESULT 34
ADR15680
ID ADR15680 standard; protein; 1078 AA.
XX
AC ADR15680;
XX
DT 04-NOV-2004 (first entry)
XX
DE Kinase 698561 hCT1827780 1, SEQ ID 73.
XX
KW Cytostatic; Cardiovascular; Neuroprotective; Nootropic; Antiparkinsonian;
KW Virucide; Cerebroprotective; Antibacterial; Fungicide; Ophthalmological;
KW Antimigraine; Analgesic; Endocrine; Tranquillizer; Hypotensive;
KW Immunosuppressive; Gene Therapy; kinase; enzyme; cancer;
KW immune-related disease; cardiovascular disease;
KW neuronal-associated disease; metabolic disorder.
XX
OS Homo sapiens.
XX
PN WO2004069154-A2.
XX
PD 19-AUG-2004.
XX
PF 28-JAN-2003; 2003WO-US002234.
XX
PR 28-JAN-2003; 2003WO-US002234.
XX
PA (GRIG/) GRIGORIEV I V.
PA (SUDA/) SUDARSANAM S.
XX
PI Grigoriev IV, Sudarsanam S;
XX
DR WPI; 2004-604329/58.
DR N-PSDB; ADR15757.
XX
PT New isolated, enriched, or purified kinase nucleic acids and
PT polypeptides, useful for diagnosing or treating kinase-related diseases
PT and conditions, e.g. cardiovascular disease, brain or neuronal-associated
PT diseases, or metabolic disorders.
XX
PS Claim 7; Fig 2; 496pp; English.
XX
CC The present invention relates to a method for detecting remote
CC polypeptide homologues, comprising analysis of conserved secondary
CC structure pattern in a protein family, and conserved active site amino
CC acid residues. The analyses are used to identify conserved residues
CC embedded into the secondary structure pattern (CRISSP), which are used to
CC detect remote homologues of the referent protein family, wherein said
CC referent protein family is the protein kinase family. The present
CC sequence is a kinase, used to illustrate the method of the invention. The
CC kinases are useful for diagnosing or treating various kinase-related
CC diseases and conditions. Diseases or disorders include cancers, immune-
CC related diseases and disorders, cardiovascular disease, brain or neuronal
CC -associated diseases, or metabolic disorders. Preferably, the diseases or
CC disorders are cancers of tissues, cancers of haematopoietic origin,
CC diseases of the central nervous system, diseases of the peripheral
CC nervous system, Alzheimer's disease, Parkinson's disease, multiple
CC sclerosis, amylotrophic lateral sclerosis, viral infections, infections
CC caused by prions, infections caused by bacteria, infections caused by
CC fungi, or ocular diseases. The disease or disorder is also migraines,
CC pain, sexual dysfunction, mood disorders, attention disorders, cognition
CC disorders, hypertension, psychotic disorders, neurological disorders,
CC dyskinesias, metabolic disorders, or organ transplant rejection.
XX
SQ Sequence 1078 AA;

```
  Query Match             36.8%;  Score 50;  DB 8;  Length 1078;
  Best Local Similarity   44.4%;  Pred. No. 1.4e+02;
  Matches   12;  Conservative    4;  Mismatches   11;  Indels    0;  Gaps    0;

Qy          2 YVFSTEMANKAAEAVLKGQVETIVSFH 28
              :|   ||  :||:||  | :|   |||
Db        469 FVRCQEMGARAAKAVESGALELSPSFH 495
```

RESULT 35
ABP66271
ID ABP66271 standard; protein; 187 AA.
XX
AC ABP66271;

```
XX
DT   19-NOV-2002  (first entry)
XX
DE   Bifidobacterium longum NCC2705 ORF amino acid sequence SEQ ID NO:1015.
XX
KW   Bifidobacterium longum NCC2705; Bifidobacterium; bacterial;
KW   antidiarrheic; antibacterial; inhibitor of Salmonella; detection;
KW   identification; lactic acid bacterium; diarrhoea; pathogenic bacteria;
KW   rotavirus; food composition; pharmaceutical composition.
XX
OS   Bifidobacterium longum.
XX
PN   EP1227152-A1.
XX
PD   31-JUL-2002.
XX
PF   30-JAN-2001; 2001EP-00102050.
XX
PR   30-JAN-2001; 2001EP-00102050.
XX
PA   (NEST ) SOC PROD NESTLE SA.
XX
DR   WPI; 2002-668397/72.
XX
PT   Novel polynucleotide comprising Bifidobacterium genome sequence useful as
PT   a probe or primer for detecting and/or identifying Bifidobacterium longum
PT   in a biological sample.
XX
PS   Claim 3; SEQ ID NO 1015; 80pp; English.
XX
CC   The present invention describes a polynucleotide (I) comprising a
CC   sequence of a Bifidobacterium genome selected from the nucleotide
CC   sequences given in ABQ81842 and ABQ81843, or a sequence exhibiting at
CC   least 90% identity or which hybridises with the sequences given in
CC   ABQ81842 and ABQ81843. Also described is a polynucleotide (II) encoding a
CC   fusion protein, comprising a sequence selected from 1097 sequences given
CC   in ABP65258 to ABP66354 ligated in frame to a polynucleotide encoding a
CC   heterologous polypeptide. (I) has antidiarrheic and antibacterial
CC   activities, and can be used as an inhibitor of Salmonella. (I) (which is
CC   a probe) is useful for the detection and/or identification of
CC   Bifidobacterium longum in a biological sample. A carrier containing the
CC   lactic acid bacterium Bifidobacterium longum NCC2705 (CNCM I-2618) can be
CC   used for preventing and/or treating diarrhoea brought about by pathogenic
CC   bacteria and/or rotavirus. The carrier is a food composition selected
CC   from milk, yogurt, curd, cheese, fermented milks, milk based fermented
CC   products, ice-creams, fermented cereal based products, milk based
CC   powders, infant formula, pet food or a pharmaceutical composition
CC   selected from tablets, liquid bacterial suspensions, dried oral
CC   supplement, wet oral supplement, dry tube feeding or wet tube feeding.
CC   (I) is useful in DNA arrays or chips to carry out analysis of the
CC   expression of the Bifidobacterium gene. ABQ81844 to ABQ81850 represent
CC   Bifidobacterium related nucleotide sequences given in the Sequence
CC   Listing from the present invention but not mentioned further within the
CC   specification. N.B. The sequence data for this patent is not represented
CC   in the printed specification but is based on sequence information
CC   supplied by the European Patent Office
XX
SQ   Sequence 187 AA;

  Query Match             36.0%;  Score 49;  DB 5;  Length 187;
  Best Local Similarity   45.2%;  Pred. No. 24;
  Matches   14;  Conservative    3;  Mismatches    4;  Indels   10;  Gaps    1;

Qy          7 EMANKAAE----------AVLKGQVETIVSF 27
              |||  |:|          ||||| | |:|:|
Db         26 EMAALASEDYRDKNPLLVAVLKGAVNTLVAF 56


RESULT 36
ABO61084
ID   ABO61084 standard; protein; 265 AA.
XX
AC   ABO61084;
XX
DT   29-JUL-2004  (first entry)
XX
DE   Klebsiella pneumoniae polypeptide seqid 7601.
XX
KW   Recombinant expression vector; transcription regulatory element;
KW   Klebsiella pneumoniae protein; antibacterial; Vaccine.
XX
OS   Klebsiella pneumoniae.
XX
PN   US6610836-B1.
XX
PD   26-AUG-2003.
XX
PF   27-JAN-2000; 2000US-00489039.
XX
PR   29-JAN-1999;   99US-0117747P.
XX
PA   (GENO-) GENOME THERAPEUTICS CORP.
XX
PI   Breton GL,  Osborne M;
XX
DR   WPI; 2003-895346/82.
DR   N-PSDB; ACH94635.
XX
PT   New nucleic acid encoding a Klebsiella pneumoniae polypeptide, useful for
PT   preparing a vaccine composition against Klebsiella pneumoniae.
XX
PS   Disclosure; SEQ ID NO 7601; 932pp; English.
XX
CC   The invention describes a new isolated nucleic acid encoding a Klebsiella
CC   pneumoniae polypeptide. Also described are: a recombinant expression
CC   vector comprising the nucleic acid, operably linked to a transcription
CC   regulatory element; and a cell comprising the recombinant expression
CC   vector. The nucleic acid is useful for preparing a vaccine composition
CC   against Klebsiella pneumoniae. This is the amino acid sequence of a
CC   Klebsiella pneumoniae polypeptide of the invention
XX
SQ   Sequence 265 AA;

  Query Match             36.0%;  Score 49;  DB 7;  Length 265;
  Best Local Similarity   45.0%;  Pred. No. 37;
  Matches    9;  Conservative    5;  Mismatches    6;  Indels    0;  Gaps    0;

Qy          8 MANKAAEAVLKGQVETIVSF 27
              |||:|  |::|  |||:  :
Db        164 MANRAPYAIMKSAVETLTRY 183


RESULT 37
ABU49676
ID   ABU49676 standard; protein; 586 AA.
XX
AC   ABU49676;
XX
DT   19-JUN-2003  (first entry)
XX
DE   Protein encoded by Prokaryotic essential gene #35203.
XX
KW   Antisense; prokaryotic essential gene; cell proliferation; drug design.
XX
OS   Vibrio cholerae.
XX
PN   WO200277183-A2.
XX
PD   03-OCT-2002.
XX
PF   21-MAR-2002; 2002WO-US009107.
XX
PR   21-MAR-2001; 2001US-00815242.
PR   06-SEP-2001; 2001US-00948993.
PR   25-OCT-2001; 2001US-0342923P.
PR   08-FEB-2002; 2002US-00072851.
```

```
PR   06-MAR-2002; 2002US-0362699P.
XX
PA   (ELIT-) ELITRA PHARM INC.
XX
PI   Wang L,  Zamudio C,  Malone C,  Haselbeck R,  Ohlsen KL,  Zyskind JW;
PI   Wall D,  Trawick JD,  Carr GJ,  Yamamoto R,  Forsyth RA,  Xu HH;
XX
DR   WPI; 2003-029926/02.
DR   N-PSDB; ACA53546.
XX
PT   New antisense nucleic acids, useful for identifying proteins or screening
PT   for homologous nucleic acids required for cellular proliferation to
PT   isolate candidate molecules for rational drug discovery programs.
XX
PS   Claim 25; SEQ ID NO 77600; 1766pp; English.
XX
CC   The invention relates to an isolated nucleic acid comprising any one of
CC   the 6213 antisense sequences given in the specification where expression
CC   of the nucleic acid inhibits proliferation of a cell. Also included are:
CC   (1) a vector comprising a promoter operably linked to the nucleic acid
CC   encoding a polypeptide whose expression is inhibited by the antisense
CC   nucleic acid; (2) a host cell containing the vector; (3) an isolated
CC   polypeptide or its fragment whose expression is inhibited by the
CC   antisense nucleic acid; (4) an antibody capable of specifically binding
CC   the polypeptide;(5) producing the polypeptide; (6) inhibiting cellular
CC   proliferation or the activity of a gene in an operon required for
CC   proliferation; (7) identifying a compound that influences the activity of
CC   the gene product or that has an activity against a biological pathway
CC   required for proliferation, or that inhibits cellular proliferation; (8)
CC   identifying a gene required for cellular proliferation or the biological
CC   pathway in which a proliferation-required gene or its gene product lies
CC   or a gene on which the test compound that inhibits proliferation of an
CC   organism acts; (9) manufacturing an antibiotic; (10) profiling a
CC   compound's activity; (11) a culture comprising strains in which the gene
CC   product is overexpressed or underexpressed; (12) determining the extent
CC   to which each of the strains is present in a culture or collection of
CC   strains; or (13) identifying the target of a compound that inhibits the
CC   proliferation of an organism. The antisense nucleic acids are useful for
CC   identifying proteins or screening for homologous nucleic acids required
CC   for cellular proliferation to isolate candidate molecules for rational
CC   drug discovery programs, or for screening homologous nucleic acids
CC   required for proliferation in cells other than S. aureus, S. typhimurium,
CC   K. pneumoniae or P. aeruginosa. The present sequence is encoded by one of
CC   the target prokaryotic essential genes. Note: The sequence data for this
CC   patent did not form part of the printed specification, but was obtained
CC   in electronic format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 586 AA;

  Query Match             36.0%;  Score 49;  DB 6;  Length 586;
  Best Local Similarity   25.0%;  Pred. No. 95;
  Matches   11;  Conservative    8;  Mismatches    7;  Indels   18;  Gaps    1;

Qy          2 YVFSTEMANKA-----------------AEAVLKGQVETIVSF 27
              |:| :::  ||                  |  ::|| ::||| |
Db          9 YIFYSQVMTKAPLLLGLVTLIGYWLLRRDATTIIKGSIKTIVGF 52


RESULT 38
ABJ25889
ID   ABJ25889 standard; protein; 984 AA.
XX
AC   ABJ25889;
XX
DT   16-APR-2003  (first entry)
XX
DE   Aspergillus fumigatus essential gene protein #547.
XX
KW   Fungicide; cytostatic; essential gene; Aspergillus fumigatus; infection;
KW   cancer; contamination; biofilm; antibody; immune response.
XX
OS   Aspergillus fumigatus.
XX
PN   WO200286090-A2.
XX
PD   31-OCT-2002.
XX
PF   23-APR-2002; 2002WO-US013142.
XX
PR   23-APR-2001; 2001US-0285697P.
PR   27-APR-2001; 2001US-0287066P.
PR   05-JUN-2001; 2001US-0295890P.
PR   09-JUL-2001; 2001US-0303899P.
PR   31-AUG-2001; 2001US-0316362P.
XX
PA   (ELIT-) ELITRA PHARM INC.
XX
PI   Jiang B,  Tishkoff D,  Zamudio C,  Eroshkin AM,  Hu W,  Lemieux SM;
XX
DR   WPI; 2003-093124/08.
XX
PT   New purified or isolated nucleic acids of essential genes of Aspergillus
PT   fumigatus, useful for treating or preventing infections by A. fumigatus,
PT   or for treating a non-infectious disease in a subject e.g. cancer.
XX
PS   Disclosure; Page; 175pp; English.
XX
CC   The invention relates to novel purified or isolated nucleic acids of
CC   essential genes of Aspergillus fumigatus. The isolated nucleic acids of
CC   the invention are used to treat or prevent infections by a pathogenic
CC   organism such as A. fumigatus, to treat a non-infectious disease in a
CC   subject (e.g. cancer), to prevent or contain contamination of an object
CC   by A. fumigatus, or to prevent or inhibit formation on a surface of a
CC   biofilm comprising A. fumigatus. The polynucleotides are useful for
CC   expressing recombinant protein for characterisation, screening or
CC   therapeutic use, as markers for host tissues in which the pathogenic
CC   organisms invade or reside, for comparing with the DNA sequence of A.
CC   fumigatus to identify duplicated genes or paralogues having the same or
CC   similar biochemical activity and/or function, for comparing with DNA
CC   sequences of other related or distant pathogenic organisms to identify
CC   potential orthologous essential or virulence genes, for selecting and
CC   making oligomers for attachment to a nucleic acid array for examination
CC   of expression patterns, for raising anti-protein antibodies, as an
CC   antigen to raise anti-DNA antibodies or to elicit another immune
CC   response, and for identifying polynucleotides encoding the other protein
CC   with which binding occurs or to identify inhibitors of the binding
CC   interaction. The polypeptides may be used to raise antibodies or to
CC   elicit immune response, as a reagent in assays designed to quantitatively
CC   determine levels of the protein in biological fluids, as a marker for
CC   host tissues in which pathogenic organism invade or reside, and to
CC   isolate correlative receptors or ligands in the case or virulence
CC   factors. This sequence represents a protein of one of the essential genes
CC   of Aspergillus fumigatus of the invention
XX
SQ   Sequence 984 AA;

  Query Match             36.0%;  Score 49;  DB 6;  Length 984;
  Best Local Similarity   52.6%;  Pred. No. 1.8e+02;
  Matches   10;  Conservative    4;  Mismatches    5;  Indels    0;  Gaps    0;

Qy          3 VFSTEMANKAAEAVLKGQV 21
              |:   ||||:| |||| ::
Db        500 VYRANMANKSAAAVLKSKL 518


RESULT 39
ABJ26489
ID   ABJ26489 standard; protein; 1058 AA.
XX
AC   ABJ26489;
XX
DT   16-APR-2003  (first entry)
XX
DE   Aspergillus fumigatus essential gene protein #1147.
XX
```

KW Fungicide; cytostatic; essential gene; Aspergillus fumigatus; infection;
KW cancer; contamination; biofilm; antibody; immune response.
XX
OS Aspergillus fumigatus.
XX
PN WO200286090-A2.
XX
PD 31-OCT-2002.
XX
PF 23-APR-2002; 2002WO-US013142.
XX
PR 23-APR-2001; 2001US-0285697P.
PR 27-APR-2001; 2001US-0287066P.
PR 05-JUN-2001; 2001US-0295890P.
PR 09-JUL-2001; 2001US-0303899P.
PR 31-AUG-2001; 2001US-0316362P.
XX
PA (ELIT-) ELITRA PHARM INC.
XX
PI Jiang B, Tishkoff D, Zamudio C, Eroshkin AM, Hu W, Lemieux SM;
XX
DR WPI; 2003-093124/08.
XX
PT New purified or isolated nucleic acids of essential genes of Aspergillus
PT fumigatus, useful for treating or preventing infections by A. fumigatus,
PT or for treating a non-infectious disease in a subject e.g. cancer.
XX
PS Disclosure; Page; 175pp; English.
XX
CC The invention relates to novel purified or isolated nucleic acids of
CC essential genes of Aspergillus fumigatus. The isolated nucleic acids of
CC the invention are used to treat or prevent infections by a pathogenic
CC organism such as A. fumigatus, to treat a non-infectious disease in a
CC subject (e.g. cancer), to prevent or contain contamination of an object
CC by A. fumigatus, or to prevent or inhibit formation on a surface of a
CC biofilm comprising A. fumigatus. The polynucleotides are useful for
CC expressing recombinant protein for characterisation, screening or
CC therapeutic use, as markers for host tissues in which the pathogenic
CC organisms invade or reside, for comparing with the DNA sequence of A.
CC fumigatus to identify duplicated genes or paralogues having the same or
CC similar biochemical activity and/or function, for comparing with DNA
CC sequences of other related or distant pathogenic organisms to identify
CC potential orthologous essential or virulence genes, for selecting and
CC making oligomers for attachment to a nucleic acid array for examination
CC of expression patterns, for raising anti-protein antibodies, as an
CC antigen to raise anti-DNA antibodies or to elicit another immune
CC response, and for identifying polynucleotides encoding the other protein
CC with which binding occurs or to identify inhibitors of the binding
CC interaction. The polypeptides may be used to raise antibodies or to
CC elicit immune response, as a reagent in assays designed to quantitatively
CC determine levels of the protein in biological fluids, as a marker for
CC host tissues in which pathogenic organism invade or reside, and to
CC isolate correlative receptors or ligands in the case or virulence
CC factors. This sequence represents a protein of one of the essential genes
CC of Aspergillus fumigatus of the invention
XX
SQ Sequence 1058 AA;

Query Match 36.0%; Score 49; DB 6; Length 1058;
Best Local Similarity 52.6%; Pred. No. 1.9e+02;
Matches 10; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

Qy 3 VFSTEMANKAAEAVLKGQV 21
|: ||||:| |||| ::
Db 500 VYRANMANKSAAAVLKSKL 518

RESULT 40
ADC50023
ID ADC50023 standard; protein; 310 AA.
XX
AC ADC50023;
XX
DT 18-DEC-2003 (first entry)
XX
DE Gene repair function associated protein-34.1.
XX
KW Gene repair function associated protein-34.1; macroprotein-51.59;
KW recombinant production; gene therapy; tumour; cancer; blood disease;
KW HIV infection; human immunodeficiency virus; immune disorder; cytostatic;
KW immunomodulator.
XX
OS Unidentified.
XX
PN CN1382717-A.
XX
PD 04-DEC-2002.
XX
PF 26-APR-2001; 2001CN-00112750.
XX
PR 26-APR-2001; 2001CN-00112750.
XX
PA (BIOW-) BIOWINDOW GENE DEV INC SHANGHAI.
XX
PI Mao Y, Xie Y;
XX
DR WPI; 2003-269479/27.
DR N-PSDB; ADC50022.
XX
PT New gene repair function associated protein-34.1, encoding
PT polynucleotide, antagonist, and recombinant production, useful for
PT treating cancer, hemopathy, HIV and immune diseases.
XX
PS Claim 1; SEQ ID NO 2; 33pp; Chinese.
XX
CC The invention relates to gene repair function associated protein-34.1
CC (ADC50023) and nucleic acids encoding it (ADC50022). The protein has a
CC molecular weight of 34.1 kD, and has 34% identity and 47% homology over a
CC 235 amino acid stretch with an Arabidopsis thaliana DNA repair protein-
CC like protein (GenBank accession number AB016875). The invention also
CC relates to a method for the recombinant production of the protein, an
CC antagonist of the protein, and the use of the protein, gene and
CC antagonist in therapeutic applications. Gene repair function associated
CC protein-34.1 can be used in the treatment of a variety of diseases such
CC as cancer, blood diseases, HIV (human immunodeficiency virus) infection
CC and immune disorders. The present sequence represents gene repair
CC function associated protein-34.1.
XX
SQ Sequence 310 AA;

Query Match 35.3%; Score 48; DB 7; Length 310;
Best Local Similarity 39.3%; Pred. No. 64;
Matches 11; Conservative 7; Mismatches 6; Indels 4; Gaps 1;

Qy 2 YVFSTEMANKAAEAVLKG----QVETIV 25
|||::| ||: : :|| :|| :|
Db 116 YVFTSEKANQEEDIPVKGSHSTKVEAVV 143

Search completed: June 8, 2005, 03:17:47
Job time : 110.125 secs